

Psychological Bulletin

CONTENTS

ARTICLES:

- Processes Affecting Scores on "Understanding of Others" and "Assumed Similarity" LEE J. CRONBACH 177
- Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores PAUL E. MEEHL AND ALBERT ROSEN 194
- Intra-Individual Response Variability DONALD W. FISKE AND LAURA RICE 217

SPECIAL REVIEW:

- Some Recent Books on Counseling and Adjustment EDWARD JOSEPH SHOREN, JR. 251

BOOK REVIEWS:

- Carmichael's Manual of Child Psychology ROGER G. BARKER 263
- Thrall, Coombs, and Davis' Decision Processes LEE J. CRONBACH 267
- Moses' The Voice of Neurosis ROY M. HAMLIN 268
- Nuttin's *Tâche Réussite et Échec. Théorie de la Conduite Humaine* MICHAEL J. ZIGLER 269
- Kornhauser, Dubin, and Ross' Industrial Conflict JEROME H. ELY 270
- Schafer's Psychoanalytic Interpretation in Rorschach Testing ROY M. HAMLIN 271
- Taylor's Dynamic and Abnormal Psychology CHARLES N. COWEN 272
- Thorpe and Schuller's Contemporary Theories of Learning LAWRENCE M. STOLUNOW 273
- Gruenberg's The Encyclopedia of Child Care and Guidance LEONARD S. KOGAN 274
- Sonnemann's Existence and Therapy ALBERT ELLIS 275
- Pennington and Berg's An Introduction to Clinical Psychology NICHOLAS HOBBS 276
- King's Psychomotor Aspects of Mental Disease: An Experimental Study JAMES D. PAGE 277
- Thelen's Dynamics of Groups at Work ALFRED MCCLUNG LEE 278

EDITORIAL NOTE 280

Published Bimonthly by the
American Psychological Association

Editorial Board (Book Reviews)

ROBERT L. TAYLOR, Editor (Statistics)

Yale College, New Haven, Conn.

Editor

Consulting Editor

LAURENCE L. CHASE,
Harvard Medical School,
Porter Hall, Cambridge,
Mass.

VICTOR C. RABIN,
University of Colorado

JOHN J. UNDERWOOD,
Yale University

WILLIAM WALLACE,
Insurance Agency
Management Association

The *Psychological Bulletin* contains the best services of research findings and articles on research methodology in psychology. The JOURNAL does not publish reports of original research or clinical observations.

Manuscripts should be sent to Wayman P. Woodworth, Department of Psychology, Teachers College, Columbia University, New York, N. Y.

Manuscripts sent to the Editor after having been reviewed will no longer appear in the JOURNAL, but in *Contemporary Psychology*, a new APA publication starting January, 1964. Manuscripts for review should be sent to E. G. Silling, Memorial Hall, Harvard University, Cambridge, Mass.

Preparation of articles for publication. Authors are strongly advised to follow the special guidelines given in the publication program of the American Psychological Association (1960, *Publications Manual*, 1960-61, Part 2), 399-449. The following instructions apply to the format and organization of the references (pp. 421-429). The format is a simplified version of that used in long reviews of books and reports. All references in double space, including the references. All manuscripts must be submitted in duplicate. Original manuscripts prepared for publication duplicate forms may be photographic or pen-and-ink copies. Authors are requested to submit a copy of the manuscript to the Editor by mail.

Books. If possible, authors are asked to contribute to articles, books, and book reviews. The editor of the JOURNAL is compelled grate to the authors of book reviews.

Correspondence, including instructions, should be sent to last issues, and change of address—should be addressed to the American Psychological Association, 1201 Connecticut Avenue, N.W., Washington 25, D. C. Changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee delivery of forwarding postage. Other claims for undelivered copies must be made within 6 months of publication.

Annual subscription: \$2.00 (Foreign \$2.50); single copies, \$1.50.

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

1201 Connecticut Avenue, N.W., Washington 25, D.C.

Printed at special rates for libraries. Under the act of October 3, 1917, authorized by the Copyright Act of 1909, and the act of February 23, 1931, authorized by the Copyright Act of 1909, the American Psychological Association, Inc., is authorized to print and distribute this publication at special rates to libraries.

Copyright 1963 by the American Psychological Association, Inc.

Psychological Bulletin

PROCESSES AFFECTING SCORES ON "UNDERSTANDING OF OTHERS" AND "ASSUMED SIMILARITY"¹

LEE J. CRONBACH

College of Education, University of Illinois

How one person judges another is a problem important for its theoretical implications and for its practical significance in group psychology, assessment, teaching, etc. Recent studies of "social perception," as this area may be termed, have been chiefly concerned with differences among perceivers either in terms of their accuracy or in terms of their tendency to view others as similar to themselves.

These studies have usually been built around a particular operation in which a Judge (*J*) "predicts" how another person (*O*) will respond. Often, for example, both persons describe themselves on a personality inventory, and *J* is then asked to fill out the inventory as he thinks *O* did. The extent to which the prediction agrees with *O*'s actual response is taken as a measure of *J*'s accuracy of social perception (or "empathy," "social sensitivity," "diagnostic competence," etc.). Scores obtained in this manner are difficult to interpret, and several investigators have reported distressingly low consistency

for them (10, 15, 32). Although this paper discusses only the perception of Others, many of the findings have relevance also to studies of "insight" where a comparison is made between the subject's self-rating and the rating given him by Others.

This paper seeks to disentangle some of the many effects which contribute to social perception scores, and to identify separately measurable components. This analysis (*a*) shows that investigators run much risk of giving psychological interpretation to mathematical artifacts when they use measures which combine the components, (*b*) directs attention to some especially interesting aspects of social perception left untouched by the usual approach, and (*c*) suggests new ideas regarding the practical use of judgments in clinic, school, and other places.

Our analysis of social perception scores may be instructive regarding research strategy generally. This area of research has developed in an ultra-operationalist manner; of late, workers have seemed content to regard "empathy" as "what empathy tests measure." The principal research activity has been correlating "empathy," so defined, with other variables. We shall show, however, that the operation involves many unsuspected sources of variation, so that scores are impure and results uninterpretable (see also 16). Studies based

¹Appreciation is expressed to Mary E. Ehart, who assisted in all stages of this paper from initial conception to final interpretation, and to Urie Bronfenbrenner and associates for helpfully providing data and for their courtesy in exchanging ideas throughout our rather similar investigations. This study was conducted under ONR Contract N6ori-07135, of which Fred E. Fiedler is now principal investigator.

on myopic operationism are largely wasted effort when the operation does not correspond to potentially meaningful constructs. Defining a measure operationally is only a preliminary to analytic studies which can refine the measure and bring it closer to the intended construct.

Our report on a specialized area of perceptual research shares much of the perspective of Postman's important general review of perception (26). His remarks are peculiarly pertinent to studies of social perception, even though he was thinking more of the "New Look" studies of perception of words and objects:

At this juncture of debate, we shall do well to pull up short a moment and reconsider the fundamental operations of our perceptual experiments, particularly as they bear on the validity of the theoretical constructs linking perception to motivation and personality. . . . Experiments have shared a common tendency which may be called the projective bias—a selective emphasis on central motivational determinants at the expense of adequate attention to the verbal and motor response dispositions of the subject and the relation of these dispositions to the dimensions of the stimulus. . . . We must then reaffirm the critical importance of a full and precise analysis of the responses as well as the stimuli which furnish the basic data of perceptual experiments (pp. 17-19).

COMPONENTS OF THE ACCURACY SCORE

In the typical experiment we have O 's self-description x_{oi} on a set of items, and a set of predictions y_{oij} by J . Error in prediction is represented by the discrepancy between x_{oi} and y_{oij} . An over-all score representing J 's ability to perceive others is obtained by averaging his squared errors over all items and all O s. The precise mathematical formulation of this Accuracy score, and the assumptions involved, are presented in the Appendix. It is important to renew here (see 8, p. 457) the warning that any index combining results from heterogeneous items presents serious

difficulties of interpretation. Whatever factors the items measure, a "global" measure combines with definite weights into a composite. Effects which operate differently on the several factors may be masked. An accuracy score based on heterogeneous items is only an exploratory procedure; where possible it should be replaced or extended by separate analyses of J 's ability to predict different qualities of O .

The Appendix shows that the usual Accuracy score is the sum of four components we shall call Elevation (E), Differential Elevation (DE), Stereotype Accuracy (SA), and Differential Accuracy (DA).

Elevation (E). The Elevation component has the form $(\bar{y}_{..j} - \bar{x}_{..})^2$. $\bar{y}_{..j}$ is the average of J 's predictions over all items and all O s; it reflects his way of using the response scale. The Elevation component is increased by any difference between J 's central tendency of responding and the central tendency of the self-descriptions, for all items and O s combined.

Differential Elevation (DE). Differential Elevation reflects how closely J 's average prediction for O corresponds to O 's central tendency of response, all items pooled and J 's central tendency of response held constant. That is, it reports J 's ability to judge deviations of the individuals' elevation from the average. We may write DE in this form:

$$DE_j^2 = \sigma_{y_{..j}}^2 + \sigma_{x_{..}}^2 - 2\sigma_{y_{..j}, x_{..}} \quad [1]$$

The variance $\sigma_{y_{..j}}^2$ expresses J 's report of how much O s will differ in elevation. This assumed dispersion in elevation will later appear as a component of the Assumed Similarity score. $\sigma_{x_{..}}^2$ is the true dispersion in elevation. The correlation $r_{y_{..j}, x_{..}}$ (to be symbolized DEr) represents J 's ability to judge which O s rate highest on the elevation scale.

In some tests, central tendency of

response (elevation) reflects insignificant response sets. In other tests, elevation has an important psychological meaning. Thus, if a high score on each item indicates morale, the correlation DEr shows how well J can judge which O s say they have the highest morale.

Stereotype Accuracy (SA). Stereotype Accuracy describes J 's ability to predict the norm for O s. It might be called "accuracy in predicting the generalized other" (3). This score depends on J 's knowledge of the relative frequency or popularity of the possible responses.

We may write:

$$SA^2 = \sigma_{g,i}^2 + \sigma_{z,i}^2 - 2\sigma_{g,i}\sigma_{z,i}r_{g,i,z,i} \quad [2]$$

Here each variance is computed over items. The variance $\sigma_{z,i}^2$ is the scatter of the actual means. SA represents ability to predict the profile of item means both as to shape and scatter. $r_{g,i,z,i}$ (Stereotype Correlation, SAr) represents accuracy in

be regarded as combining variances with a correlation term. The correlation (DAr) measures ability to judge which O s have highest scores on the item, when the score is taken as a deviation from O s' mean. There is one such correlation for each item.

Removal of the E and SA components reduces all data to deviations from the group mean, or from the predicted mean. The DE component examines ability to recognize individual differences in the first centroid factor underlying the items. The DA component pools all remaining factors. There is nothing to prevent estimating further factors among the items by factor analysis, and determining DA for each factor separately. This type of cluster score on major factors appears preferable theoretically to the simpler index, $\sum_i DA_i$, and more reliable than the DA_i taken separately (8).

Implications. Seven aspects of J 's performance have been separated:

- | | |
|--|-----------------------------------|
| 1. Elevation component (E) | |
| 2. Assumed dispersion in Elevation | } Differential Elevation (DE) |
| 3. Elevation correlation (DEr) | |
| 4. Predicted variation in item means | } Stereotype Accuracy (SA) |
| 5. Stereotype correlation (SAr) | |
| 9. Assumed dispersion on any item (Elevation held constant) | } Differential Accuracy (DA) |
| 7. Differential correlation (DAr) | |

judging mean profile shape, i.e., the order of item difficulties.

Differential Accuracy (DA). Differential Accuracy reflects ability to predict differences between O s on any item. This component is averaged over items. As the Appendix indicates, this component too may

The components are not necessarily uncorrelated. Change in any component alters the Accuracy score.

Surely these aspects of social perception do not all reflect the same trait. A Judge who happens to use the same region of the response scale as other persons (Elevation is small)

need not have superior insight. Judging which items have the highest mean seems to require acquaintance with the norms of the group; but a person might possess such knowledge and yet be unable to differentiate accurately between individuals (16).

At best, failure to separate these components makes interpretation ambiguous. Chowdhry and Newcomb (5) requested group members to predict what percentage of their group would agree with each of many attitude statements. Ability to make this prediction was judged by a difference score, and this score was found to correlate significantly with leadership status. This score, however, combines Elevation and Stereotype Accuracy; until the components are separately measured we cannot rule out the possibility that leaders simply used the correct range of the scale more often than nonleaders. This, in turn, might reflect willingness (or unwillingness) to use extreme percentages rather than any other subtle perceptiveness of specific attitudes. That such effects do occur is shown by Lorge and Diamond, who required judges to estimate what proportion of *O*s would pass ability test items. They found that poor judges were greatly helped simply by being told the difficulty of a few specimen items. "Apparently the difference between 'poorest,' 'mediocre,' and 'best' judges is that the 'best' judges have some experiential referent for the per cent of the population that can pass an item. Giving such referents to the 'poorest' and 'mediocre' judges . . . leads to a significant reorientation of such judgments" (19, p. 33). When judges responded only to the items, the best judges had a mean *SA_r* of .73 and the poorest, .56. After information on averages for items was given, the same judges had mean correlations of .77 and .73. The difficulty

encountered in interpreting the Chowdhry-Newcomb study does not arise in Talland's study of the same problem (33) where subjects were asked to predict what ranks will be assigned to certain stimuli. In ranking, elevation and dispersion are the same for everyone, and therefore the scores depend only on *SA_r*.

Failure to identify the components of the Accuracy score can lead to artifactual correlations. Only a few of the many examples in the literature need be cited. Norman and Ainsworth (24) report a large number of correlations between Accuracy ("empathy") and Assumed Similarity ("projection"). Since the Accuracy score contains Assumed Similarity components, the two scores would necessarily overlap even if both sets of responses are determined strictly by chance. The correlations have no psychological meaning. Dymond (11) reported that persons with high Accuracy are also most easily judged. But a person who uses the scale in a typical manner will have a small Elevation component, hence better Accuracy; and other persons will have smaller Elevation errors in judging him, simply because of this typicality. This would happen even if the Judge predicted his responses without ever meeting him! Perhaps social psychologists should take what comfort they can from Bertrand Russell's remark that physicists "have not yet reached the point where they can distinguish between facts about relativity and mathematical operations which may have nothing to do therewith."

COMPONENTS OF THE ASSUMED SIMILARITY SCORE

The *J* is said to "assume similarity" between himself and *O* if *J*'s prediction for *O* differs little from *J*'s self-description. One might study assumed similarity with respect to

each O separately, but we shall give attention to the tendency of J to assume similarity over O s in general. The formula for this AS score is given in the Appendix. AS is sometimes interpreted as "projection" or "identification" (29). As indicated in Equation 5a of the Appendix, the AS score divides into the components Assumed Elevation (AE), Assumed Self-Typicality (AST), and two Assumed Dispersions (ADE , ADI).

Assumed Similarity in Elevation (AE). The first component, Assumed Similarity in Elevation, takes the form $(\bar{y}_{..j} - \bar{x}_{..j})^2$. It measures J 's tendency to assume that O s have the same average response as he does. This component would be important if items are polarized so that a high score on each represents good adjustment or some other interpretable quality; the score then shows whether J regards the average O as similar to himself in this central dimension.

Assumed Dispersions (ADE , ADI). A second component is $\sigma_{j..}^2$, the Assumed Dispersion in Elevation (ADE). Another is the Assumed Dispersion on specific items after differences in Elevation are removed (ADI). ADI closely resembles Gage's concept of "rigidity" or "adherence to stereotype" in prediction (14, p. 16; 15). These dispersions have already been encountered as components of ACC (see Equation 1 above and Equation 4a in Appendix). We shall refer to them as Assumed Dispersion in Elevation (ADE) and Assumed Dispersion Within Items (ADI), respectively.

Assumed Self-Typicality (AST). The remaining component measures the discrepancy between J 's perception of the average O and his self-description. This component tells whether J regards his own profile as typical in shape. Or, we might say, this component shows the simi-

larity of J 's self-perception to his implicit stereotype of O s (Elevation held constant). We follow Gage (14, p. 17) in calling this Assumed Self-Typicality (AST).

Of the four AS components, only AST divides into separate variance and correlational terms, as shown in Equation 6a of the Appendix. The correlation represents the similarity between J 's self-description and the average profile, ignoring differences in elevation and scatter. We call it the Self-Typicality Correlation (STR).

IMPROVEMENT OF PREDICTIONS

Insofar as our mathematical model is an acceptable approximation to real problems, we can reason mathematically to determine how judgments of O s may be improved. The conditions which make errors of prediction as small as possible are stated fully in the Appendix. The most significant principle takes this form: Accuracy is improved as σ_y approaches $r_{xy}\sigma_x$. That is to say, the variation in predictions should never exceed the variation in true responses, and should ordinarily be much smaller.

This principle indicates that there is an optimal degree of differentiation in making judgments. If J can make accurate judgments as to the relative location of O s on a continuum, then he is wise to make σ_y as large as σ_x —never larger. But if he is forced to base his judgment on inadequate cues or if the available personality theory and situational knowledge do not permit trustworthy inference, then he should treat people as if they were very nearly alike. The person who attempts to differentiate individuals on inadequate data introduces error even when the inferences have validity greater than chance.

The variation of J 's predictions indicates how much he differentiates. For example, a teacher estimating

IQ's in a class might spread them from 90 to 110 or from 70 to 130. We would expect the judge who perceives greater differences to apply more sharply differentiated treatments to the various persons. A person who knows that the expected σ for IQ's is 16 might try to predict so that his estimates would have this σ ; but unless he is a perfect judge, this is unwise. He will have smaller errors if his predicted σ is less than 16—how much less depending on the correlational accuracy of his predictions.

If two diagnosticians can each judge some trait with correlational validity .40, the one who differentiates strongly (i.e., makes extreme statements) will make far more serious absolute errors than the one who differentiates moderately. Indeed, the person who makes extreme differentiations based on a validity of .40 may have larger errors than a judge who has zero correlational validity but gives the same estimate for everyone. This contradicts the view that judgment is always improved by taking into account additional valid information.

Implications regarding clinical judgment. Clinical judgments are frequently regarded as undependable, because of research tending to show that clinicians weight predictor variables inappropriately. Thus Sarbin showed that counselors given a great amount of information predicted grade averages no more accurately than did a regression formula, one reason being that the counselors gave the *ACE* test excessive weight (27). A similar problem of weighting is involved in attaining the ideal dispersion of estimates. The regression equation for two uncorrelated predictors in standard form is:

$$x_{1.2} = r_{12}x_2 + r_{13}x_3 + w_e x_e \quad [3]$$

The w_e in the last term is a weight for error, and x_e is an individual's estimated error score. We do not usually write this term because x_e is zero and the term vanishes. Since $\sigma_{1.2} = \sqrt{1 - w_e^2} = R\sigma_1$, the statistical prediction formula gives the optimum degree of differentiation, using the proper w_e . But the clinician who combines variables 2 and 3 with the best relative weights may still obtain an inaccurate estimate if he employs the wrong w_e , and so makes $\sigma_{1.2}$ too large.

There is evidence, both from Sarbin's study and from similar recent work by R. S. Melton (20, 21) that counselors do overdifferentiate. In Sarbin's study, σ_e for grade-point averages was .88 (for both sexes, computed by this writer from Sarbin's report). The statistical predictions had a σ of .47; the clinical predictions, .57. In unpublished data supplied by Melton, σ_e was .59, but σ_e (for predictions) was .60. The optimum, $r\sigma_e$, would presumably have been near .35.

Sarbin was puzzled by the finding that $\sigma_e < \sigma_e$, and he discusses this as a "disadvantage" (28, p. 599). His remarks resemble those of others on the so-called "central tendency of judgment," which has hitherto been regarded as a source of inaccuracy in social perception (1, p. 521). But estimates from a regression formula have a lower variance than observed results just because weight must be assigned to error. Sarbin's clinicians made some, but insufficient, allowance for their error. Melton's clinicians made no allowance for error. Evidently the fault of the clinician is too little "central tendency of judgment."

Melton's study combines all sources of error into a single measure of absolute error. He reports that counselors make greater error than a statistical predictor, and that this

remains true even when the clinicians are given an actuarial table as a guide. Decomposing his accuracy score into components yields further knowledge about the predictive process. From the data he has supplied us, we learn that the group given the actuarial table makes no error in estimating mean GPA, while the control group overestimates by $.65\sigma$. Both groups differentiate far more than the optimum (and this effect appears to be intensified with the actuarial table). From the data at hand, we do not know whether correlation accuracy (D_{Ar}) is higher or lower for Melton's counselors than for a statistical prediction, nor how the actuarial table altered this aspect of their judgment.

Systematic errors such as over-optimism and overdifferentiation may be corrected fairly easily. It is important for studies of clinical judgment to measure these errors as separate components, and for clinicians to train themselves to avoid these errors.

Implications for teaching. Recognizing an optimum degree of differentiation makes it necessary to re-examine and qualify statements commonly made in training teachers, to the effect that every pupil has his own pattern and the teacher must fit methods to that pattern, not treat the pupil in terms of the statistical average. The writer has himself expressed such views, but it now appears that the teacher who is poorly informed regarding the unique patterns of his pupils should probably treat them by a standard pattern of instruction, carefully fitted to the typical pupil. Modifying plans drastically on the basis of limited diagnostic information may do harm. Differentiation is harmful if the extent of adaptation or differentiation exceeds the amount justified by the accuracy of social perception.

Teachers may properly modify treatments considerably when they are well able to judge individual differences. Differences in arithmetic achievement they might judge quite accurately; if so, they could profitably provide quite different assignments for different individuals. But if it is hard to judge creative potential in art, say, or the ultimate vocational goal of a ninth-grader, then it is a great mistake to differentiate treatment to fit perceived differences.

ILLUSTRATIVE ANALYSIS OF CORNELL DATA

To illustrate our system of analysis, we use data kindly provided by Bronfenbrenner and Dempsey. The data were gathered at Cornell University primarily for pilot analyses such as ours. We shall deal with eight subjects and eight items. The eight subjects were candidates for employment as interviewers. Each person interviewed each of the seven others. In each interview, each man was to obtain information about his partner. Following the interview, each person stated his own reaction to 19 items (of which we use only eight) and predicted what his partner would say. One item is: To what extent did you feel at ease during the interview? ____a. very much ____b. a good bit ____c. only slightly ____d. not at all. The respective responses are scored 1-2-3-4.

Completion of the design provides seven self-descriptions and seven predictions by each man (also seven for each man). We have taken two simplifying steps which might be illegitimate for purposes other than demonstration. We use the average of O 's responses over all seven interviews as his true response, x_{oi} , discarding information on O 's variation from interview to interview. Secondly, we treat J 's self-description as a "perfectly accurate prediction of

TABLE 1
ACCURACY SCORES OF EIGHT JUDGES DIVIDED INTO COMPONENTS

| Judge | ACC ^a | Elevation (E) | Differential Elevation (DE) | Stereotype Component (SA) | Differential Component (DA) | DE Contains | | SA Contains | | DA Contains | |
|----------|------------------|------------------|-----------------------------------|---------------------------------|-----------------------------------|----------------|----------------------------------|----------------|----------------------------------|------------------|------------------------------------|
| | | | | | | $\sigma_{\%J}$ | $r_{\%J, \sigma_{\%J}}$ (DER) | $\sigma_{\%J}$ | $r_{\%J, \sigma_{\%J}}$ (SAR) | $\sigma_{\%J}^2$ | $r_{\%J, \sigma_{\%J}^2}$ (DAR) |
| 1 | 2.32 | .00 | .69 | .24 | 1.40 | .24 | .18 | .47 | .93 | .12 | .28 |
| 2 | 5.32 | .05 | 1.71 | 1.34 | 2.24 | .38 | -.14 | .17 | .37 | .17 | .12 |
| 3 | 4.60 | .10 | .23 | .50 | 3.77 | .15 | .63 | .47 | .85 | .39 | .08 |
| 4 | 5.33 | .26 | 1.35 | .91 | 2.81 | .35 | .00 | .21 | .67 | .25 | .23 |
| 5 | 4.37 | .98 | .79 | .28 | 2.33 | .27 | .18 | .45 | .91 | .16 | -.04 |
| 6 | 4.06 | 1.22 | .76 | .65 | 1.45 | .18 | -.20 | .54 | .86 | .11 | .26 |
| 7 | 3.75 | .20 | .73 | 1.27 | 1.56 | .14 | -.44 | .15 | .44 | .17 | .38 |
| 8 | 3.78 | .62 | .41 | 1.28 | 2.06 | .29 | .62 | .76 | .92 | .27 | .35 |
| Mean | 4.19 | .35 | .83 | .81 | 2.20 | .25† | .10 | .40† | .74 | .20 | .21 |
| Variance | .83 | .20 | .20 | .18 | .56 | .01 | .13 | .04 | .04 | .01 | .02 |

* Averaged over items.

† The values of $\sigma_{\%J}$, .25 and .40, may be compared to these respective values of $\sigma_{\%J}$, the true variation: .22, .44.

himself." By this device, we deal at all times with eight Js and eight Os, and the criterion is made the same for every person.

Accuracy Scores for Eight Persons

Table 1 presents the ACC score for each person, and his score on each component. These and subsequent results are illustrative, and not a proper basis for generalization.

Relation of differentiation to accuracy. As expected, any component decreases as the predicted standard deviation ($\sigma_{\%J}$) approaches the product of the related correlation with the actual standard deviation ($\sigma_{\%J}$). Consider, for example, the results on SA and its constituents. Person 8 is an excellent Judge, according to his SAR of .92. But he expects too much variation in the item means (.76 compared to an actual σ of .44). As a consequence No. 8 has a poor SA score despite his excellent ability to discriminate between items. The best SA scores are earned by No. 1 and No. 5, who have high correlations and who predicted variance close to the actual variance. Compare also DE and DER of No. 3 and No. 8. These persons have the same DER, but No. 3 underestimates the variation in elevation, and No. 8 overesti-

mates it. As expected, No. 3 earns the better DE score.

The judges consistently overdifferentiate. The optimal $\sigma_{\%J}$ would be about .025 in the DE column, and .30 in SA; but the actual mean values are .25 and .40. Since $\sigma_{\%J}^2 = .12$, it is clear that $\sigma_{\%J}^2$ is also generally larger than the optimum.

Reliabilities and intercorrelations. Internal consistency was studied for the various components, but the results based on eight cases need not be reported. One finding, however, is notable. Differential Accuracy was strikingly consistent over items: a coefficient of .73 was obtained by analysis of variance. That is to say, some predictors were consistently good over all items, others consistently poor. But when we examine the components of DA, we find that Assumed Dispersion Within Items is consistent over items (.79), and DAR, the measure of accuracy in differentiating, is not (.18). In this sample, Differential Accuracy shows reliability only because some persons have consistent sets to differentiate. Stone and Leavitt (32) likewise find very low consistency (-.07 to .30) of accuracy scores in predicting different children, but a median consistency of .63 between two predictions for

TABLE 2
ASSUMED SIMILARITY SCORES DIVIDED INTO COMPONENTS

| Judge | AS ² | Assumed Elevation (AE) | Assumed Dispersion in Elevation (ADE) | Assumed Self- Typicality (AST)* | Self- Typicality Correlation (STr) | Assumed Dispersion Within Items (ADI) |
|----------|-----------------|------------------------------|--|--|---|--|
| 1 | 1.89 | .14 | .46 | .34 | .94 | .97 |
| 2 | 3.78 | .35 | 1.15 | .95 | .60 | 1.36 |
| 3 | 4.63 | .06 | .18 | 1.30 | .90 | 3.09 |
| 4 | 3.45 | .16 | .98 | .30 | .61 | 2.01 |
| 5 | 2.21 | .01 | .59 | .37 | .89 | 1.25 |
| 6 | 1.54 | .12 | .26 | .28 | .94 | .87 |
| 7 | 2.41 | .00 | .15 | .88 | .72 | 1.37 |
| 8 | 4.15 | .12 | .66 | 1.16 | .89 | 2.20 |
| Mean | 3.01 | .12 | .55 | .70 | .81 | 1.64 |
| Variance | 1.14 | .23 | .12 | .16 | .02 | .49 |

* A composite of *STr* and $\sigma_{g,ij}$ of Table 1. See Equation 6a of Appendix.

the same child. They trace the latter consistency to consistent favorable sets toward a given child, and to assumed similarity. All results to date lead us to doubt whether accuracy in differentiating personalities of others can be reliably measured. Where reliable variance is found, it seems to result from some constant mental set.

In Table 1 we note that No. 1 is consistently superior on various components of Accuracy and No. 4 is consistently inferior. But No. 7, the best predictor as judged by *DAr*, is the poorest on *DEr* and next to poorest on *SAr*. With only eight cases, meaningful correlations cannot be obtained.

Future studies of predictive accuracy should measure the components separately, preferably using two independent sets of items and *Os*. Such measurement will permit accurate determination of reliabilities of components, of the relation between the components, and of their relation, if any, to external criteria. Ideally, items would be organized into clusters to permit study of predictions on separate traits. Only after such research can we decide

how many components within the over-all Accuracy score presently used are important, and which unwanted components must be suppressed by appropriate design of tests and scoring keys.

Assumed Similarity Scores for Eight Persons

In Table 2, the Assumed Similarity scores are divided into components. The relatively large variance of *ADI* indicates that it has great influence on individual differences in over-all *AS*.

AE correlates .81 with *ADE*, and *AST* correlates .97 with *ADI*. In these data, the tendency to differentiate among *Os* is accompanied by a tendency to differentiate the average *O* from oneself. This result is partly an artifact, resulting from using each person's self-description as one of his "predictions." Even allowing for this, our correlations suggest separating only two components of *AE*: Assumed Similarity in Elevation (*AE*+*ADE*) and Assumed Similarity in Pattern (*AST*+*ADI*). The correlation between these is only .21. Further evidence is required, however, to establish definitely how to

divide Assumed Similarity. An earlier study (9) suggests strongly that Assumed Similarity is a general mental set, almost independent of the psychological content of the items. A global index may therefore be satisfactory for this score.

THE JUDGE'S "IMPLICIT PERSONALITY THEORY"

We turn now to an aspect of social perception data which may prove to be particularly significant. When a Judge describes or makes predictions for a large number of *O*s, these predictions define a distribution of points in the variate space. This distribution may be regarded as a description of the generalized *O*, representing *J*'s view of both central tendency and individual differences. The *J*'s generalized perception may be an important indicator of his expectations regarding *O*s. We shall discuss the general significance of this perceptual system before tracing its effect on social perception scores.

The *J*'s distribution is to be examined in terms of the means, variances, and covariances of the predictions. The mean may be regarded as *J*'s stereotype; if the mean *O* in his descriptions is "hostile," for example, this may be highly significant. The variance or assumed dispersion on a variate indicates *J*'s tendency to differentiate along that dimension. The covariance is interpreted as indicating the relation *J* expects to find among variates. A given *J* may customarily report the same persons as high on both "quietness" and "shyness," for instance; or on both "ambition" and "selfishness." These aspects of the distribution reveal *J*'s view of *O*s and the connotation of personality traits for him. We suggest that these means, variances, and covariances describe *J*'s implicit theory of personality.

The expectations *J* has of *O*s con-

stitute his view of personality, and one may hypothesize that they direct his responses to *O*s. G. A. Kelly (18) argues that each person forms "personal constructs" by means of which he differentiates situations confronting him, including other persons. The "personal constructs" would be, in our model, the dimensions along which *J* differentiates strongly. As Steiner (31, p. 349) notes, it is particularly important to investigate the covariation of his constructs. Osgood (25) suggests studying the semantic equivalence of stimuli by testing whether they are used similarly. Our method is quite like his, determining as it does what traits *J* uses to describe the same persons. While Steiner asks a person directly what traits he expects to be associated, we suggest looking at the covariation found among the predictions y_{jt} . Such implicit relations are not subject to deliberate distortion and can reveal associations and norms of which *J* himself is unaware.

An illustrative case. This concept can be illustrated by using a small portion of the Bronfenbrenner-Dempsey data. The *J* predicted responses of eight persons (including himself) on these questions:

1. In general, how openly did you express your feelings and emotions during the interview?
2. How much interest did you feel in the other man as a person?
3. How much were you aware of how he was feeling?
4. How much opportunity did you give him to interview you?
5. How much important information were you able to get about him?
6. To what extent did you feel at ease during the interview?
7. To what extent did you succeed in establishing a good interviewing relationship?
8. To what extent did you feel like the person being interviewed rather than the person doing the interviewing?

The matrix of covariances for Judge 3, a poor predictor, was fac-

TABLE 3
SUBSTANTIAL FACTORS IN JUDGE 3'S COVARIANCE MATRIX*

| Item | Factor I "Pressure" | Factor II "Exchange of Inform- ation" | Factor III "Rapport" | <i>h²</i> | Variance <i>s_{total}²</i> | Mean Prediction <i>y_a</i> |
|-------------------------|------------------------|--|-------------------------|----------------------|--|--|
| 1 | <i>2.55</i> | -.68 | -.62 | 7.35 | 7.58 | 2.39 |
| 2 | 1.32 | -.94 | -.16 | 2.65 | 3.77 | 1.80 |
| 3 | -.76 | .16 | .82 | 1.28 | 1.81 | 1.70 |
| 4 | .46 | <i>1.35</i> | -.07 | 2.04 | 2.34 | 1.63 |
| 5 | .68 | <i>1.17</i> | .16 | 1.86 | 2.16 | 1.59 |
| 6 | <i>-1.92</i> | -.76 | <i>1.28</i> | 5.90 | 6.42 | 1.79 |
| 7 | -.88 | -.30 | .63 | 1.26 | 2.14 | 1.43 |
| 8 | <i>2.50</i> | -.08 | .40 | 6.43 | 6.43 | 2.93 |
| Per cent of variance | 62 | 16 | 10 | 88 | | |
| Cumulative per cent | 62 | 70 | 88 | | | |

* Largest loadings in each column in italics.

tored by a pivotal method intended to yield interpretable factors. Table 3 shows the loadings on three factors, and also item means and variances. The means for Judge 3 show no striking features, especially when considered in relation to the true means presented in Table 4. The variances indicate that No. 3 regards others as fairly uniform in their awareness of him (item 3), and as varying especially in openness, ease, and feeling of dominance (items 1, 6, 8). No confidence can be placed in factors based on eight measures, but we would otherwise interpret Factor I as representing a feeling of being under pressure. It is notable that No. 3 regards those persons who are most open (item 1) as being least at ease (item 6). Factor II shows a link between items 4 and 5, getting and giving information. Factor III is indistinct. It is notable that items 6 and 7 are correlated; a "good interviewing relation" is perceived by No. 3 as one where the *interviewer* is at ease! Such a finding regarding No. 3's perception, if better substantiated, might have

much diagnostic importance.

Relevant prior studies. The literature contains many studies of correlation between ratings which bear on the perceiver's frame of reference. Reports of halo effect suggest the existence of a strong general good-bad factor. These studies have not examined raters separately. Newcomb (23) showed that there were substantial correlations among ratings ($r = .49$) even where direct behavioral observations on the same qualities showed a mean correlation of only .14. "The close relation . . .," says Newcomb, "may be presumed to spring from logical presuppositions in the minds of the raters" (p. 288). Steiner (31) found evidence that ethnocentric individuals see others in black and white terms, the "good," "strong" traits going together. In our language, their covariance matrix is loaded with one factor, while non-authoritarians use many factors and do not emphasize the general evaluative dimension. Soskin (30) showed that halo effect and the stereotype profile varied as a function of the data given the assessor and that

TABLE 4
FACTORS IN THE CRITERION COVARIANCE MATRIX DETERMINED BY PIVOTAL METHOD*

| Item | Factor I "Openness" | Factor II "Receptiveness" | Factor III "Passivity" | λ^2 | Variance $\sigma_{\epsilon_i}^2$ | Mean \bar{x}_i |
|----------------------|------------------------|------------------------------|---------------------------|-------------|-------------------------------------|---------------------|
| 1 | <i>2.06</i> | -.40 | -.30 | 4.49 | 4.57 | 2.05 |
| 2 | .14 | .94 | .01 | .88 | 1.03 | 1.93 |
| 3 | .09 | .96 | -.11 | .94 | 1.16 | 2.04 |
| 4 | .86 | .40 | .30 | 1.00 | 1.07 | 1.80 |
| 5 | .22 | .55 | -.44 | .54 | .67 | 2.05 |
| 6 | .47 | .60 | <i>1.16</i> | 1.93 | 1.42 | 1.57 |
| 7 | -.10 | .86 | .09 | .76 | .97 | 1.64 |
| 8 | -.05 | -.27 | .85 | .80 | 2.45 | 3.09 |
| Per cent of variance | 40 | 27 | 18 | 85 | | |
| Cumulative per cent | 40 | 67 | 85 | | | |

* Largest loadings in each column in italics.

ratings of peers are distributed differently from ratings by professional assessors.

A striking recent study by Jones (17) compares the ratings given by authoritarian (A) and nonauthoritarian (NA) groups to Others regarding whom carefully controlled information had been given. He finds many types of differences including a tendency of the NA's to respond to personal qualities of the Other, whereas A's seem to differentiate less among leaders. The data are analyzed to show, in effect, which traits perceived in the Other vary with the Other's *democracy*. The groups agree in associating democracy with *sensitive to Other, generous, adaptable, warm*. The A's associate *democrat* also with *unambitious, poor officer, undependable, hard to figure out, acts without thinking, rebellious*; while the NA's associate these qualities with *autocrat*. An attempt was made also to find correlates of *forceful*. The two groups showed little difference in covariances, associating *forceful* with *natural leader, ambitious, uses his head*, etc. At least two other studies show differences in

the perceptual reference frames of groups. Wickman's well-known study (34) showed that teachers expected different traits to correlate with mental health than did mental hygienists. Moore (22) performed a factor analysis of ratings given noncommissioned officers by their subordinates, and also of ratings given by their superiors. The factor patterns differed. For instance, superiors coupled *leadership* with *eagerness* and *responsibility*, but the subordinates linked *leadership* with *intelligence* and *skill*.

None of these studies of groups examines the perceptual space by which an individual describes personality, but the evidence supports the belief that important individual differences exist. Our proposed analysis of the covariance matrix can be applied to the matrix based on mean ratings given by a group of raters. This should give more complete information regarding group differences in implicit meanings than the methods used in the studies cited. The emphasis in recent studies has been to consider correlations between items as a meaningful phenomenon, re-

lated to psychologically interesting qualities of the rater. This may be contrasted to the view in the earliest studies such as Newcomb's, where such correspondences were regarded solely as an annoying interference or so-called "logical error" in rating. It also contrasts with many recent studies which concentrate on the interaction between Perceiver and Other, failing to inquire about elements associated with the Perceiver alone.

The analysis of the implicit meanings of various dimensions for the Perceiver may be used in several ways. If, in teacher training, one aim is to modify the way in which teachers interpret behavior, these changes in viewpoint should be reflected in changes of the perceptual distribution. For example, if teachers naively regard *quietness* as associated with *adjustment*, yet experts regard quietness as unrelated or even negatively related to adjustment in children, then the aim of training is to reduce or reverse the correlation found in the teachers' responses. Because our technique examines implicit interpretations, it should be especially useful for evaluation. Another application of the method is in industrial rating. The rater's distribution, when rating many applicants on many traits, indicates what differences he pays attention to and how he interprets the traits he is supposed to rate. In analyzing one rater in this way, we find that he regards *creative* and *inquiring* as nearly independent of *intelligent*, and their independent contribution is negatively (!) correlated with his final recommendation as to hiring. So far as we can determine, previous studies have pooled all raters before studying trait intercorrelations;³ the

study of idiosyncratic rating patterns should lead to important suggestions for training raters or improving scales. Finally, in view of our interpretation of the perceptual distribution as an implicit personality theory, special interest would attach to studies of ratings given by clinical psychologists or psychiatrists of different schools, who might be presumed to hold different theories.

Effect on accuracy scores. The *J*'s distribution of *O*s has been interpreted here as a standing system of meanings which delimits the space within which he locates *O*s. It is obvious that any such delimitation would affect social perception scores. Discrepancies between mean and actual mean lower Stereotype Accuracy, and Accuracy declines if perceived variance (*ADE*, *ADI*) departs from an optimal value. The correlational effects are a bit less easy to perceive.

Correlations describe the shape of the distribution of *O*s. If traits 1 and 2 are uncorrelated, then x_{o1} , x_{o2} will have a roughly circular joint distribution. If a *J* regards 1 and 2 as correlated, his perceived distribution of y_{o1} , y_{o2} will be elliptical. Perceived variance along the dimension 1+2 will become greater than in the true responses, and Accuracy will suffer. We can view the example in another way. Suppose the Judge predicts variate 1 perfectly but believes that variates 1 and 2 correlate 1.00—then he must have substantial error in predicting variate 2. He can predict 2 accurately only if he perceives the covariance of 1 with 2 accurately.

Data reported by Crow (10, p. 86) show this phenomenon clearly. He asked *J*s to predict what would be the first word missed by a patient on

³ Especially clear evidence on this point is provided in a recent dissertation by Walker. (Walker, W. B. An investigation of the effectiveness of communication between psy-

chologists and sales executives through personnel audit reports. Unpublished doctor's dissertation, Western Reserve Univer., 1955.)

a vocabulary test and what would be the highest level attained (called tasks D1 and D2). The correlation of *Js'* Accuracy on D1 with Accuracy on D2 was positive and significant for five of ten patients, but negative and significant on two patients. Judges tended to *expect* a correlation between D1 and D2, and they were accurate on those patients where the two scores were similar. Where the scores were dissimilar, *Js* could not be accurate on both predictions. There was a rank correlation of .97 (over *Os*) between Accuracy, and consistency of *O's* performance.

We can provide further illustration from the Cornell data. The covariance between items in self-descriptions was factored, with the results shown in Table 4. This pattern is different from that of No. 3 (Table 3) in several respects. Notably, No. 3 overdifferentiates on all items. The first factor for No. 3 lumps openness and lack of receptiveness; these variables are divided among two factors in the criterion. In the criterion, being at ease (item 6) is positively related to openness. It is especially interesting that "feeling like the person being interviewed" is, for the group as a whole, positively correlated with being at ease; but for No. 3 these items are negatively correlated. When his expectancy is so discrepant from the facts, it is not surprising that No. 3 has poor accuracy.

RECOMMENDATIONS

Studies of perception may be concerned either with constant processes or with variable processes. When social perception is regarded (as in 1, pp. 499-548) as a process of interpreting the expressive cues *O* presents, or of empathizing with him, the search is clearly for a variable process. The concept of an "intuitive" perception of *Os* which underlies much of the relevant research

implies that *J* is reacting to the particular *O* as a stimulus, and ignores the fact that the perceptual response also depends on stereotypes in *J's* mind (cf. Cattell, 5). We have seen that the measures currently used are affected by both constant and reactive processes, and therefore cannot serve well to investigate either. As Crow states:

The difficulty stems from failure to recognize that two meanings of predictive accuracy are involved. The use of the correlation scoring method [either $r_{x_{ij}y_{ij}}$ or $r_{x'_{ij}y'_{ij}}$] defines predictive accuracy as the ability to vary one's predictions as the actual situation varies. The difference score method defines predictive accuracy as the ability to approximate the actual situation. By the difference score method a subject is penalized for a systematic error in estimation of the magnitude of the actual situation. By the correlation method the subject is not so penalized. Conversely, a subject is penalized by the correlation method if, although he has approximated the actual situation, his predictions do not vary concomitantly with the actual scores. Each of these scoring methods has its advantages and disadvantages. The choice of which technique to use will depend on the purpose for which a study is conducted, although a second basis for choice depends upon the empirical relationship between two procedures (10, p. 57).

An argument can be presented for concentrating attention on constant processes, taking up interactions between *J* and *O* only after the constant processes characteristic of *J* are dependably measured. Constant processes in the perceiver have potentially great importance because they affect all his acts of perception. Individual differences in constant processes need to be measured dependably so that their influence can be discounted in studies of variable processes. Moreover, *identifying constant errors should permit training to eliminate such biases*; this may be the most effective way to improve the social perception of leaders, teachers, and diagnosticians.

Not all constant processes are of theoretical importance. We ven-

ture to suggest which components of social perception measures deserve attention, recognizing that the ultimate importance of the components depends on whether they relate to important criteria.

1. To some extent, the Elevation components (*E*, *DE*, *DEr*) reflect whether *J* interprets the words defining the scale in the same manner as *Os* do. It appears relatively unfruitful, therefore, as a source of information on his perception of *Os* (7; 8, p. 463). It should be separately measured where it is believed to have psychological significance, and otherwise eliminated from consideration. This is consistent with Postman's view:

In experiments concerned with the determinants of perceptual selectivity, the contribution of verbal and motor response habits must be specifically evaluated and wherever possible held constant. The effects of the independent variables can then be evaluated against an empirical baseline defined by the response habits of the subjects (26, p. 26).

2. The Assumed Similarity measures reflect a general orientation toward Others. Perhaps the tendency to differentiate which these indices measure is a reaction shown only in the testing situation. But the fact that significant behavioral correlates have been found for Assumed Similarity (2, 4, 12, 13, 29) suggests that this is a generalized mental set influencing both test and nontest behavior. Investigators would do well, however, to consider Postman's conclusion that response dispositions can be established unambiguously only if they are measured by more than one type of response (26, p. 27).

Components of Assumed Similarity include Assumed Dispersion in Elevation, Assumed Dispersion over Items, Assumed Similarity in Elevation, and Assumed Self-Typicality. Further research is required to determine whether these should be measured separately or combined.

3. Stereotype Accuracy expresses how closely *J*'s implicit picture of the generalized Other agrees with reality. Differences of this sort are probably important. Attention should be given to the nature of *J*'s errors, as well as to the over-all magnitude of the component.

4. The *J*'s perceptual space, studied as a whole, includes not only information on his stereotype and his assumed dispersion, but also on the way in which he organizes the field of personality. This type of constant cognitive process appears to be a most important area for research.

5. The Differential Elevation Correlation and the Differential Accuracy Correlation are measures of *J*'s sensitivity to individual differences. These measures reflect his ability to interpret expressive behavior, or his ability in differential diagnosis. These are the only processes included in present measures of social perception which depend on *J*'s sensitivity to the particular *O*. The reliability of measures of this variable process has not been encouraging. But those who wish to study "empathy" or "social sensitivity" as it has usually been conceptualized should extract these correlational components from their measures.

Social perception research has been dominated by simple, operationally defined measures. Our analysis has shown that any such measure may combine and thereby conceal important variables, or may depend heavily on unwanted components. Only by careful subdivision of global measures can an investigator hope to know what he is dealing with. Our analysis makes especially clear that the investigator of social perception must develop more explicit theory regarding the constructs he intends to study, so that he can reduce his measures to the genuinely relevant components.

APPENDIX

To simplify this paper for the reader, we have placed our detailed mathematical argument here. In our notation, x_{oi} is the self-description of Other o on item i ($i=a, b, c, \dots, k$). y_{oi} is Judge j 's description of o on i . We employ the customary notation for means: \bar{x}_o indicates the average of x_{oi} over all items, \bar{x}_i the average over Others, and $\bar{x}_{..}$ the grand mean over Others and items. Averages on y are defined similarly. $x_{oi}' = x_{oi} - \bar{x}_o - \bar{x}_i + \bar{x}_{..}$; that is, the score x_{oi} is transformed as a deviation from both item mean and Other mean. y_{oi}' is defined similarly.

Error in prediction may be measured by $|y_{oi} - x_{oi}|$. We shall, however, use the squared difference. This formula is easier to treat mathematically than the absolute difference, and will ordinarily give similar results. When all items are in a Yes-No form, so that the error on any prediction is 1 or 0, the two methods give identical results. Our measure has the important property of being invariant under orthogonal rotation of axes (8).

The Accuracy with which J perceives all Others is defined by

$$ACC_j^2 = \frac{1}{kN} \sum_o \sum_i (y_{oi} - x_{oi})^2. \quad [1a]$$

The following identity may be written:

$$\begin{aligned} y_{oi} - x_{oi} &= (y_{..i} - \bar{y}_{..}) - (\bar{x}_o - \bar{x}_{..}) \\ &\quad + [(y_{oi} - y_{..i}) - (\bar{x}_o - \bar{x}_{..})] \\ &\quad + [(y_{..i} - \bar{y}_{..}) - (\bar{x}_i - \bar{x}_{..})] \\ &\quad + (y_{oi}' - x_{oi}'). \end{aligned} \quad [2a]$$

When we square and sum, cross-products drop out and we have the resolution of ACC_j^2 into components:

$$\begin{aligned} ACC_j^2 &= \frac{1}{kN} \sum_o ACC_{oi} = (y_{..i} - \bar{y}_{..})^2 \\ &\quad + \frac{1}{N} \sum_o [(y_{oi} - y_{..i}) - (\bar{x}_o - \bar{x}_{..})]^2 \\ &\quad + \frac{1}{k} \sum_i [(y_{..i} - \bar{y}_{..}) - (\bar{x}_i - \bar{x}_{..})]^2 \\ &\quad + \frac{1}{kN} \sum_o \sum_i (y_{oi}' - x_{oi}')^2. \end{aligned} \quad [3a]$$

In order, these components are called E , DE , SA , and DA (squares being ignored except in equations). Each of the three latter terms may be rewritten as the variance of a difference. Equations 1 and 2 in the text indicate this form for DE and SA . We also have

$$DA_{ij}^2 = \sigma_{y'_{oi}}^2 + \sigma_{x'_{oi}}^2 - 2\sigma_{y'_{oi}x'_{oi}} = \sigma_{y'_{oi}}^2 + \sigma_{x'_{oi}}^2 - 2r_{y'_{oi}x'_{oi}} \sigma_{y'_{oi}} \sigma_{x'_{oi}}. \quad [4a]$$

DA_{ij}^2 , averaged over items yields DA_j^2 . Each variance in the formula is taken over Others.

Some investigators have preferred to compute $DA_{oj}^2 = \sum_i (x_{oi}' - y_{oi}')^2$. Summed over Others, this also yields DA_j^2 . Subdivided, DA_{oj}^2 would yield a variance over items for the Other, and a "Q correlation" over items comparing predicted deviations for the Other with the actual deviations. This method of organizing the data is not recommended, because the correlations are critically dependent on the factorial content of the items employed and on the direction chosen as representing a high score on the item. In personality data, this direction is frequently arbitrary.

Assumed similarity is also defined in terms of a sum of squared differences:

$$\begin{aligned} AS_j^2 &= \frac{1}{kN} \sum_o \sum_i (y_{oi} - x_{oi})^2 = (y_{..i} - \bar{y}_{..})^2 \\ &\quad + \sigma_{\bar{x}_o}^2 \\ &\quad + \frac{1}{k} \sum_i [(y_{..i} - \bar{y}_{..}) - (\bar{x}_i - \bar{x}_{..})]^2 \\ &\quad + \frac{1}{k} \sum_i \sigma_{y'_{oi}}^2. \end{aligned} \quad [5a]$$

The components, in order, are AE , ADE , AST , and ADI . Only AST can be rewritten as a variance of differences:

$$AST^2 = \sigma_{y_{..i}}^2 + \sigma_{x_{..i}}^2 - 2\sigma_{y_{..i}x_{..i}} = \sigma_{y_{..i}}^2 + \sigma_{x_{..i}}^2 - 2r_{y_{..i}x_{..i}} \sigma_{y_{..i}} \sigma_{x_{..i}}. \quad [6a]$$

This correlation is referred to as the Self-Typicality correlation (STr).

We assume that the goodness of predictions can be evaluated by the mean square error. Taking the derivative of each component of ACC_j^2 , and setting that derivative equal to zero, we find that ACC becomes smaller and therefore prediction improves, when:

a. J has a typical response set (E approaches zero).

b. $\sigma_{y_{..i}}^2$ approaches $r_{y_{..i}x_{..i}}^2 \sigma_{x_{..i}}^2$. Here the variance is over items. This means that $\sigma_{y_{..i}}^2$ should not exceed $\sigma_{x_{..i}}^2$, and should be near zero if the Stereotype correlation is low. If this correlation is low, the more J differentiates among items, the poorer is his Accuracy.

c. $\sigma_{y'_{oi}}^2$ approaches $r_{y'_{oi}x'_{oi}}^2 \sigma_{x'_{oi}}^2$, the variance being over Others. This means that $\sigma_{y'_{oi}}^2$ should not exceed $\sigma_{x'_{oi}}^2$, and should be near zero if the Differential correlation is low. This principle holds for accuracy of prediction on any single item, and for predicting elevation.

REFERENCES

1. ALLPORT, G. W. *Personality: a psychological interpretation*. New York: Holt, 1937.
2. BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, 48, 61-66.

3. BRONFENBRENNER, U., & HARDING, J. Components and correlates of social sensitivity. Paper read at Amer. Psychol. Ass., New York, September, 1954.
4. CASS, LORETTA K. Parent-child relationships and delinquency. *J. abnorm. soc. Psychol.*, 1952, **47**, 101-104.
5. CATTELL, R. B. Measurement versus intuition in applied psychology. *J. Pers.*, 1937, **6**, 114-131.
6. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51-57.
7. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3-31.
8. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456-473.
9. CRONBACH, L. J., HARTMANN, W., & EHART, MARY E. *Investigation of the character and properties of assumed similarity measures*. Urbana, Ill: Bureau of Research and Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 7, Contract N6ori-07135.)
10. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
11. DYMOND, ROSALIND F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343-350.
12. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101-107.
13. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381-388.
14. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
15. GAGE, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22**, 128-141.
16. GAGE, N. L., & CRONBACH, L. J. *Conceptual and methodological problems in interpersonal perception*. Urbana, Ill.: Bureau of Educational Research, Univer. of Illinois, 1954 (Mimeo.)
17. JONES, E. E. Authoritarianism as a determinant of first-impression formation. *J. Pers.*, 1954, **23**, 107-127.
18. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, in press.
19. LORGE, I., & DIAMOND, LORRAINE K. The value of information to good and poor judges of item difficulty. *Educ. psychol. Measmt.*, 1954, **14**, 29-33.
20. MELTON, R. S. A comparison of clinical and actuarial methods of prediction with an assessment of the relative accuracy of different clinicians. Unpublished doctor's dissertation, Univer. of Minnesota, 1953.
21. MELTON, R. S. A study of the relative accuracy of counselor judgments and actuarial predictions. *Amer. Psychologist*, 1954, **9**, 429-430. (Abstract)
22. MOORE, J. V. Factor analytic comparisons of superior and subordinate ratings of the same NCO supervisors. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1953, No. 53-24.
23. NEWCOMB, T. M. An experiment designed to test the validity of a rating technique. *J. educ. Psychol.*, 1937, **22**, 279-289.
24. NORMAN, R. D., & AINSWORTH, PATRICIA. The relationships among projection, empathy, reality, and adjustment, operationally defined. *J. consult. Psychol.*, 1954, **18**, 53-58.
25. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197-237.
26. POSTMAN, L. Perception, motivation, and behavior. *J. Pers.*, 1953, **22**, 17-31.
27. SARBIN, T. R. The relative accuracy of clinical and statistical predictions of academic success. Unpublished doctor's dissertation, Ohio State Univer., 1941.
28. SARBIN, T. R. A contribution to the study of actuarial and individual methods of prediction. *Amer. J. Sociol.*, 1943, **48**, 593-602.
29. SOPCHAK, A. L. Parental "identification" and "tendency toward disorders" as measured by the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.*, 1952, **47**, 159-165.
30. SOSKIN, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, **10**, 107-114.
31. STEINER, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol.*, 1954, **49**, 349-354.
32. STONE, G. C., & LEAVITT, G. S. Generality of accuracy in perceiving standard persons. Paper read at Midwest Psychol. Ass., Columbus, Ohio, May, 1954.
33. TALLAND, G. A. The assessment of group opinion by leaders, and their influence on its formation. *J. abnorm. soc. Psychol.*, 1954, **49**, 431-434.
34. WICKMAN, E. K. *Children's behavior and teachers' attitudes*. New York: Commonwealth Fund, 1928.

Received May 5, 1954.

ANTECEDENT PROBABILITY AND THE EFFICIENCY OF PSYCHOMETRIC SIGNS, PATTERNS, OR CUTTING SCORES

PAUL E. MEEHL

University of Minnesota

AND ALBERT ROSEN

VA Hospital, Minneapolis, and University of Minnesota¹

In clinical practice, psychologists frequently participate in the making of vital decisions concerning the classification, treatment, prognosis, and disposition of individuals. In their attempts to increase the number of correct classifications and predictions, psychologists have developed and applied many psychometric devices, such as patterns of test responses as well as cutting scores for scales, indices, and sign lists. Since diagnostic and prognostic statements can often be made with a high degree of accuracy purely on the basis of actuarial or experience tables (referred to hereinafter as *base rates*), a psychometric device, to be efficient, must make possible a greater number of correct decisions than could be made in terms of the base rates alone.

The efficiency of the great majority of psychometric devices reported in the clinical psychology literature is difficult or impossible to evaluate for the following reasons:

a. Base rates are virtually never reported. It is, therefore, difficult to determine whether or not a given device results in a greater number of correct decisions than would be possible solely on the basis of the rates from previous experience. When,

however, the base rates can be estimated, the reported claims of efficiency of psychometric instruments are often seen to be without foundation.

b. In most reports, the distribution data provided are insufficient for the evaluation of the probable efficiency of the device in other settings where the base rates are markedly different. Moreover, the samples are almost always too small for the determination of optimal cutting lines for various decisions.

c. Most psychometric devices are reported without cross-validation data. If a psychometric instrument is applied solely to the criterion groups from which it was developed, its reported validity and efficiency are likely to be spuriously high, especially if the criterion groups are small.

d. There is often a lack of clarity concerning the type of population in which a psychometric device can be effectively applied.

e. Results are frequently reported only in terms of significance tests for differences between groups rather than in terms of the number of correct decisions for individuals within the groups.

The purposes of this paper are to examine current methodology in studies of predictive and concurrent validity (1), and to present some methods for the evaluation of the efficiency of psychometric devices as well as for the improvement in the interpretations made from such devices. Actual studies reported in the

¹ From the Neuropsychiatric Service, VA Hospital, Minneapolis, Minnesota, and the Divisions of Psychiatry and Clinical Psychology of the University of Minnesota Medical School. The senior author carried on his part of this work in connection with his appointment to the Minnesota Center for the Philosophy of Science.

literature will be used for illustration wherever possible. It should be emphasized that these particular illustrative studies of common practices were chosen simply because they contained more complete data than are commonly reported, and were available in fairly recent publications.

IMPORTANCE OF BASE RATES

Danielson and Clark (4) have reported on the construction and application of a personality inventory which was devised for use in military induction stations as an aid in detecting those men who would not complete basic training because of psychiatric disability or AWOL recidivism. One serious defect in their article is that it reports cutting lines which have not been cross validated. Danielson and Clark state that inductees were administered the Fort Ord Inventory within two days after induction into the Army, and that all of these men were allowed to undergo basic training regardless of their test scores.

Two samples (among others) of these inductees were selected for the study of predictive validity: (a) A group of 415 men who had made a good adjustment (Good Adjustment Group), and (b) a group of 89 men who were unable to complete basic training and who were sufficiently disturbed to warrant a recommendation for discharge by a psychiatrist (Poor Adjustment Group). The authors state that "the most important task of a test designed to screen out misfits is the detection of the (latter) group" (4, p. 139). The authors found that their most effective scale for this differentiation picked up, at a given cutting point, 55% of the Poor Adjustment Group (valid positives) and 19% of the Good Adjustment Group (false positives). The overlap between these two groups

would undoubtedly have been greater if the cutting line had been cross validated on a random sample from the *entire population* of inductees, but for the purposes of the present discussion, let us assume that the results were obtained from cross-validation groups. There is no mention of the percentage of all inductees who fall into the Poor Adjustment Group, but a rough estimate will be adequate for the present discussion. Suppose that in their population of soldiers, as many as 5% make a poor adjustment and 95% make a good adjustment. The results for 10,000 cases would be as depicted in Table 1.

TABLE 1
NUMBER OF INDUCTEES IN THE POOR ADJUSTMENT AND GOOD ADJUSTMENT GROUPS DETECTED BY A SCREENING INVENTORY
(55% valid positives; 19% false positives)

| Predicted Adjustment | Actual Adjustment | | | | Total Pre- dicted |
|-------------------------|-------------------|-----|-------|-----|-------------------------|
| | Poor | | Good | | |
| | No. | % | No. | % | |
| Poor | 275 | 55 | 1,805 | 19 | 2,080 |
| Good | 225 | 45 | 7,695 | 81 | 7,920 |
| Total actual | 500 | 100 | 9,500 | 100 | 10,000 |

Efficiency in detecting poor adjustment cases. The efficiency of the scale can be evaluated in several ways. From the data in Table 1 it can be seen that if the cutting line given by the authors were used at Fort Ord, the scale could not be used directly to "screen out misfits." If all those predicted by the scale to make a poor adjustment were screened out, the number of false positives would be extremely high. Among the 10,000 potential inductees, 2080 would be predicted to make a poor adjustment. Of these 2080, only 275, or 13%, would actually make a poor adjustment, whereas the decisions

for 1805 men, or 87% of those screened out, would be incorrect.

Efficiency in prediction for all cases. If a prediction were made for every man on the basis of the cutting line given for the test, 275+7695, or 7970, out of 10,000 decisions would be correct. Without the test, however, every man would be predicted to make a good adjustment, and 9500 of the predictions would be correct. Thus, use of the test has yielded a drop from 95% to 79.7% in the total number of correct decisions.

Efficiency in detecting good adjustment cases. There is one kind of decision in which the Inventory can improve on the base rates, however. If only those men are accepted who are predicted by the Inventory to make a good adjustment, 7920 will be selected, and the outcome of 7695 of the 7920, or 97%, will be predicted correctly. This is a 2% increase in hits among predictions of "success." The decision as to whether or not the scale improves on the base rates sufficiently to warrant its use will depend on the cost of administering the testing program, the administrative feasibility of rejecting 21% of the men who passed the psychiatric screening, the cost to the Army of training the 225 maladaptive recruits, and the intangible human costs involved in psychiatric breakdown.

Populations to which the scale is applied. In the evaluation of the efficiency of any psychometric instrument, careful consideration must be given to the types of populations to which the device is to be applied. Danielson and Clark have stated that "since the final decision as to disposition is made by the psychiatrist, the test should be classified as a screening adjunct" (4, p. 138). This statement needs clarification, however, for the efficiency of the scale can vary markedly according

to the different ways in which it might be used as an adjunct.

It will be noted that the test was administered to men who were already in the Army, and not to men being examined for induction. The reported validation data apply, therefore, specifically to the population of *recent inductees*. The results might have been somewhat different if the population tested consisted of *potential inductees*. For the sake of illustration, however, let us assume that there is no difference in the test results of the two populations.

An induction station psychiatrist can use the scale cutting score in one or more of the following ways, i.e., he can apply the scale results to a variety of populations. (a) The psychiatrist's final decision to accept or reject a potential inductee may be based on both the test score and his usual interview procedure. The population to which the test scores are applied is, therefore, *potential inductees interviewed by the usual procedures for whom no decision was made*. (b) He may evaluate the potential inductee according to his usual procedures, and then consult the test score *only if* the tentative decision is to reject. That is, a decision to accept is final. The population to which the test scores are applied is *potential inductees tentatively rejected by the usual interview procedures*. (c) An alternative procedure is for the psychiatrist to consult the test score *only if* the tentative decision is to accept, the population being *potential inductees tentatively accepted by the usual interview procedures*. The decision to reject is final. (d) Probably the commonest proposal for the use of tests as screening adjuncts is that the more skilled and costly psychiatric evaluation should be made only upon the test positives, i.e., inductees classified by the test as good

risks are not interviewed, or are subjected only to a very short and superficial interview. Here the population is *all potential inductees*, the test being used to make either a *final* decision to "accept" or a decision to "examine."

Among these different procedures, how is the psychiatrist to achieve maximum effectiveness in using the test as an adjunct? There is no answer to this question from the available data, but it can be stated definitely that the data reported by Danielson and Clark apply only to the third procedure described above. The test results are based on a selected group of men *accepted* for induction and not on a random sample of potential inductees. If the scale is used in any other way than the third procedure mentioned above, the results may be considerably inferior to those reported, and, thus, to the use of the base rates without the test.²

The principles discussed thus far, although illustrated by a single study, can be generalized to any study of predictive or concurrent validity. It can be seen that many considerations are involved in determining the efficiency of a scale at a given cutting score, especially the base rates of the subclasses within the population to which the psychometric device is to be applied. In a subsequent portion of this paper, methods will be presented for determining cutting points for maximizing the efficiency of the different types of decisions which are made with psychometric devices.

Another study will be utilized to illustrate the importance of an explicit statement of the base rates of popu-

lation subgroups to be tested with a given device. Employing an interesting configural approach, Thiesen (18) discovered five Rorschach patterns, each of which differentiated well between 60 schizophrenic adult patients and a sample of 157 gainfully employed adults. The best differentiator, considering individual patterns or number of patterns, was Pattern A, which was found in 20% of the patients' records and in only .6% of the records of normals. Thiesen concludes that if these patterns stand the test of cross validation, they might have "clinical usefulness" in early detection of a schizophrenic process or as an aid to determining the gravity of an initial psychotic episode (18, p. 369). If by "clinical usefulness" is meant efficiency in a clinic or hospital for the diagnosis of schizophrenia, it is necessary to demonstrate that the patterns differentiate a higher percentage of schizophrenic patients from *other diagnostic groups* than could be correctly classified without any test at all, i.e., solely on the basis of the rates of various diagnoses in any given hospital. If a test is to be used in differential diagnosis among psychiatric patients, evidence of its efficiency for this function cannot be established solely on the basis of discrimination of diagnostic groups from normals. If by "clinical usefulness" Thiesen means that his data indicate that the patterns might be used to detect an early schizophrenic process among nonhospitalized gainfully employed adults, he would do better to discard his patterns and use the base rates, as can be seen from the following data.

Taulbee and Sisson (17) cross validated Thiesen's patterns on schizophrenic patient and normal samples, and found that Pattern A was the best discriminator. Among patients,

² Goodman (8) has discussed this same problem with reference to the supplementary use of an index for the prediction of parole violation.

8.1% demonstrated this pattern and among normals, none had this pattern. There are approximately 60 million gainfully employed adults in this country, and it has been estimated that the rate of schizophrenia in the general population is approximately .85% (2, p. 558). The results for Pattern A among a population of 10,000 gainfully employed adults would be as shown in Table 2. In order to detect 7 schizophrenics, it would be necessary to test 10,000 individuals.

TABLE 2

NUMBER OF PERSONS CLASSIFIED AS SCHIZOPHRENIC AND NORMAL BY A TEST PATTERN AMONG A POPULATION OF GAINFULLY EMPLOYED ADULTS

(8.1% valid positives; 0.0% false positives)

| Classification by Test | Criterion Classification | | | | Total Classified by Test |
|------------------------|--------------------------|------|--------|-----|--------------------------|
| | Schizophrenia | | Normal | | |
| | No. | % | No. | % | |
| Schizophrenia | 7 | 8.1 | 0 | 0 | 7 |
| Normal | 78 | 91.9 | 9,915 | 100 | 9,993 |
| Total in class | 81 | 100 | 9,915 | 100 | 10,000 |

In the Neurology service of a hospital a psychometric scale is used which is designed to differentiate between patients with psychogenic and organic low back pain (9). At a given cutting point, this scale was found to classify each group with approximately 70% effectiveness upon cross validation, i.e., 70% of cases with no organic findings scored above an optimal cutting score, and 70% of surgically verified organic cases scored below this line. Assume that 90% of all patients in the Neurology service with a primary complaint of low back pain are in fact "organic." Without any scale at all the psychol-

ogist can say every case is organic, and be right 90% of the time. With the scale the results would be as shown in Section A of Table 3. Of 10 psychogenic cases, 7 score above the line; of 90 organic cases, 63 score below the cutting line. If every case above the line is called psychogenic, only 7 of 34 will be classified correctly or about 21%. Nobody wants to be right only one out of five times in this type of situation, so that it is obvious that it would be imprudent to call a patient psychogenic on the basis of this scale. Radically different results occur in prediction for cases below the cutting line. Of 66 cases 63, or 95%, are correctly classified as organic. Now the psychologist has increased his diagnostic hits from 90 to 95% on the condition that he labels only cases falling below the line, and ignores the 34% scoring above the line.

TABLE 3

NUMBER OF PATIENTS CLASSIFIED AS PSYCHOGENIC AND ORGANIC ON A LOW BACK PAIN SCALE WHICH CLASSIFIES CORRECTLY 70% OF PSYCHOGENIC AND ORGANIC CASES

| Classification by Scale | Actual Diagnosis | | Total Classified by Scale |
|---|------------------|---------|---------------------------------|
| | Psycho- genic | Organic | |
| <i>A. Base Rates in Population Tested: 90% Organic; 10% Psychogenic</i> | | | |
| Psychogenic | 7 | 27 | 34 |
| Organic | 3 | 63 | 66 |
| Total diagnosed | 10 | 90 | 100 |
| <i>B. Base Rates in Population Tested: 90% Psychogenic; 10% Organic</i> | | | |
| Psychogenic | 63 | 3 | 66 |
| Organic | 27 | 7 | 34 |
| Total diagnosed | 90 | 10 | 100 |

In actual practice, the psychologist may not, and most likely will not,

test every low back pain case. Probably those referred for testing will be a select group, i.e., those who the neurologist believes are psychogenic because neurological findings are minimal or absent. This fact changes the population from "all patients in Neurology with a primary complaint of low back pain," to "all patients in Neurology with a primary complaint of low back pain *who are referred for testing*." Suppose that a study of past diagnoses indicated that of patients with minimal or absent findings, 90% were diagnosed as psychogenic and 10% as organic. Section B of Table 3 gives an entirely different picture of the effectiveness of the low back pain scale, and new limitations on interpretation are necessary. Now the scale correctly classifies 95% of all cases above the line as psychogenic (63 of 66), and is correct in only 21% of all cases below the line (7 of 34). In this practical situation the psychologist would be wise to refrain from interpreting a low score.

From the above illustrations it can be seen that the psychologist in interpreting a test and in evaluating its effectiveness must be very much aware of the population and its subclasses and the base rates of the behavior or event with which he is dealing at any given time.

It may be objected that no clinician relies on just one scale but would diagnose on the basis of a configuration of impressions from several tests, clinical data and history. We must, therefore, emphasize that the preceding single-scale examples were presented for simplicity only, but that the main point is not dependent upon this "atomism." *Any complex configurational procedure in any number of variables, psychometric or otherwise, eventuates in a decision.* Those decisions have a certain objective suc-

cess rate in criterion case identification; and for present purposes we simply treat the decision function, whatever its components and complexity may be, as a single variable. It should be remembered that the literature does not present us with cross-validated methods having hit rates much above those we have chosen as examples, regardless of how complex or configural the methods used. So that even if the clinician approximates an extremely complex configural function "in his head" before classifying the patient, for purposes of the present problem this complex function is treated as the scale. In connection with the more general "philosophy" of clinical decision making see Bross (3) and Meehl (12).

APPLICATIONS OF BAYES' THEOREM

Many readers will recognize the preceding numerical examples as essentially involving a principle of elementary probability theory, the so-called "Bayes' Theorem." While it has come in for some opprobrium on account of its connection with certain pre-Fisherian fallacies in statistical inference, as an algebraic statement the theorem has, of course, nothing intrinsically wrong with it and it does apply in the present case. One form of it may be stated as follows:

If there are k antecedent conditions under which an event of a given kind may occur, these conditions having the antecedent probabilities P_1, P_2, \dots, P_k of being realized, and the probability of the event upon each of them is $p_1, p_2, p_3, \dots, p_k$; then, given that the event is observed to occur, the probability that it arose on the basis of a specified one, say j , of the antecedent conditions is given by

$$P_{1(s)} = \frac{P_1 p_1}{\sum_{i=1}^k P_i p_i}$$

The usual illustration is the case of drawing marbles from an urn. Suppose we have two urns, and the urn-selection procedure is such that the probability of our choosing the first urn is 1/10 and the second 9/10. Assume that 70% of the marbles in the first urn are black, and 40% of those in the second urn are black. I now (blindfolded) "choose" an urn and then, from it, I choose a marble. The marble turns out to be black. What is the probability that I drew from the first urn?

$$P_1 = .10 \quad P_2 = .90$$

$$p_1 = .70 \quad p_2 = .40$$

Then

$$P_{1(s)} = \frac{(.10)(.70)}{(.10)(.70) + (.90)(.40)} = .163.$$

If I make a practice of inferring under such circumstances that an observed black marble arose from the first urn, I shall be correct in such judgments, in the long run, only 16.3% of the time. Note, however, that the "test item" or "sign" *black marble* is correctly "scored" in favor of Urn No. 1, since there is a 30% difference in black marble rate between it and Urn No. 2. But this considerable disparity in symptom rate is overcome by the very low base rate ("antecedent probability of choosing from the first urn"), so that inference to first-urn origin of black marbles will actually be wrong some 84 times in 100. In the clinical analogue, the urns are identified with the subpopulations of patients to be discriminated (their antecedent probabilities being equated to their base rates in the population to be examined), and the black marbles are test results of a

certain ("positive") kind. The proportion of black marbles in one urn is the valid positive rate, and in the other is the false positive rate. Inspection and suitable manipulations of the formula for the common two-category case, viz.,

$$P_{(s)} = \frac{P p_1}{P p_1 + Q p_2}$$

$P_{d(s)}$ = Probability that an individual is diseased, given that his observed test score is positive

P = Base rate of actual positives in the population examined

$$P + Q = 1$$

p_1 = Proportion of diseased identified by test ("valid positive" rate)

$$q_1 = 1 - p_1$$

p_2 = Proportion of nondiseased misidentified by test as being diseased ("false positive" rate)

$$q_2 = 1 - p_2$$

yields several useful statements. Note that in what follows we are operating entirely with exact population parameter values; i.e., sampling errors are not responsible for the dangers and restrictions set forth. See Table 4.

1. In order for a positive diagnostic assertion to be "more likely true than false," the ratio of the positive to the negative base rates in the examined population must exceed the ratio of the false positive rate to the valid positive rate. That is,

$$\frac{P}{Q} > \frac{p_2}{p_1}$$

If this condition is not met, the attribution of pathology on the basis of the test is more probably in error than correct, *even though the sign being used is valid* (i.e., $p_1 \neq p_2$).

TABLE 4
DEFINITION OF SYMBOLS

| Diagnosis from Test | Actual Diagnosis | |
|-----------------------------|--|--|
| | Positive | Negative |
| Positive | p_1 Valid positive rate (Proportion of positives called positive) | p_2 False positive rate (Proportion of negatives called positive) |
| Negative | q_1 False negative rate (Proportion of positives called negative) | q_2 Valid negative rate (Proportion of negatives called negative) |
| Total with actual diagnosis | $p_1 + q_1 = 1.0$ (Total positives) | $p_2 + q_2 = 1.0$ (Total negatives) |

Note.—For simplicity, the term "diagnosis" is used to denote the classification of any kind of pathology, behavior, or event being studied, or to denote "outcome" if a test is used for prediction. Since horizontal addition (e.g., $p_1 + p_2$) is meaningless in ignorance of the base rates, there is no symbol or marginal total for these sums. All values are parameter values.

Example: If a certain cutting score identifies 80% of patients with organic brain damage (high scores being indicative of damage) but is also exceeded by 15% of the nondamaged sent for evaluation, in order for the psychometric decision "brain damage present" to be more often true than false, the ratio of actually brain-damaged to nondamaged cases among all seen for testing must be at least one to five (.19).

Piotrowski has recommended that the presence of 5 or more Rorschach signs among 10 "organic" signs is an efficient indicator of brain damage. Dorken and Kral (5), in cross validating Piotrowski's index, found that 63% of organics and 30% of a mixed, nonorganic, psychiatric patient group had Rorschachs with 5 or more signs. Thus, our estimate of $p_2/p_1 = .30/.63 = .48$, and in order for the decision "brain damage present" to be correct more than one-half the

time, the proportion of positives (P) in a given population must exceed .33 (i.e., $P/Q > .33/.67$). Since few clinical populations requiring this clinical decision would have such a high rate of brain damage, especially among psychiatric patients, the particular cutting score advocated by Piotrowski will produce an excessive number of false positives, and the positive diagnosis will be more often wrong than right. Inasmuch as the base rates for any given behavior or pathology differ from one clinical setting to another, *an inflexible cutting score should not be advocated for any psychometric device*. This statement applies generally—thus, to indices recommended for such diverse purposes as the classification or detection of deterioration, specific symptoms, "traits," neuroticism, sexual aberration, dissimulation, suicide risk, and the like. When P is small, it may be advisable to explore the possibility of dealing with a restricted population within which the base rate of the attribute being tested is higher. This approach is discussed in an article by Rosen (14) on the detection of suicidal patients in which it is suggested that an attempt might be made to apply an index to subpopulations with higher suicide rates.

2. If the base rates are equal, the probability of a positive diagnosis being correct is the ratio of valid positive rate to the sum of valid and false positive rates. That is,

$$p_{d(o)} = \frac{p_1}{p_1 + p_2} \quad \text{if } P = Q = \frac{1}{2}.$$

Example: If our population is evenly divided between neurotic and psychotic patients the condition for being "probably right" in diagnosing psychosis by a certain method is simply that the psychotics exhibit the pattern in question more fre-

quently than the neurotics. This is the intuitively obvious special case; it is often misgeneralized to justify use of the test in those cases where base-rate asymmetry ($P \neq Q$) counteracts the ($p_1 - p_2$) discrepancy, leading to the paradoxical consequence that *deciding on the basis of more information can actually worsen the chances of a correct decision*. The apparent absurdity of such an idea has often misled psychologists into behaving as though the establishment of "validity" or "discrimination," i.e., that $p_1 \neq p_2$, indicates that a procedure should be used in decision making.

Example: A certain test is used to select those who will continue in outpatient psychotherapy (positives). It correctly identifies 75% of these good cases but the same cutting score picks up 40% of the poor risks who subsequently terminate against advice. Suppose that in the past experience of the clinic 50% of the patients terminated therapy prematurely. Correct selection of patients can be made with the given cutting score on the test 65% of the time, since $p_1/(p_1 + p_2) = .75/(.75 + .40) = .65$. It can be seen that the efficiency of the test would be exaggerated if the base rate for continuation in therapy were actually .70, but the efficiency were evaluated solely on the basis of a research study containing equal groups of continuers and noncontinuers, i.e., if it were assumed that $P = .50$.

3. In order for the hits in the entire population which is under consideration to be increased by use of the test, the base rate of the more numerous class (called here positive) must be less than the ratio of the valid negative rate to the sum of valid negative and false negative rates. That is, unless

$$P < \frac{q_2}{q_1 + q_2},$$

the making of decisions on the basis of the test will have an adverse effect. An alternative expression is that $(P/Q) < (q_2/q_1)$ when $P > Q$, i.e., the ratio of the larger to the smaller class must be less than the ratio of the valid negative rate to the false negative rate. When $P < Q$, the conditions for the test to improve upon the base rates are:

$$Q < \frac{p_1}{p_1 + p_2}$$

and

$$\frac{Q}{P} < \frac{p_1}{p_2}.$$

Rotter, Rafferty, and Lotsof (15) have reported the scores on a sentence completion test for a group of 33 "maladjusted" and 33 "adjusted" girls. They report that the use of a specified cutting score (not cross validated) will result in the correct classification of 85% of the maladjusted girls and the incorrect classification of only 15% of the adjusted girls. It is impossible to evaluate adequately the efficiency of the test unless one knows the base rates of maladjustment (P) and adjustment (Q) for the population of high school girls, although there would be general agreement that $Q > P$. Since $p_1/(p_1 + p_2) = .85/(.85 + .15) = .85$, the overall hits in diagnosis with the test will not improve on classification based solely on the base rates unless the proportion of adjusted girls is less than .85. Because the reported effectiveness of the test is spuriously high, the proportion of adjusted girls would no doubt have to be considerably less than .85. Unless there is good reason to believe that

the base rates are similar from one setting to another, it is impossible to determine the efficiency of a test such as Rotter's when the criterion is based on ratings unless one replicates his research, including the criterion ratings, with a representative sample of each new population.

4. In altering a sign, improving a scale, or shifting a cutting score, the increment in valid positives per increment in valid positive *rate* is proportional to the positive base rate; and analogously, the increment in valid negatives per increment in valid negative *rate* is proportional to the negative base rate. That is, if we alter a sign the net improvement in over-all hit rate is

$$H'_T - H_T = \Delta p_1 P + \Delta q_2 Q,$$

where H_T = original proportion of hits (over-all) and H'_T = new proportion of hits (over-all).

5. A corollary of this is that altering a sign or shifting a cut will improve our decision making if, and only if, the ratio of *improvement* Δp_1 in valid positive rate to *worsening* Δp_2 in false negative rate exceeds the ratio of actual negatives to positives in the population.

$$\frac{\Delta p_1}{\Delta p_2} > \frac{Q}{P}.$$

Example: Suppose we improve the intrinsic validity of a certain "schizophrenic index" so that it now detects 20% more schizophrenics than it formerly did, at the expense of only a 5% increase in the false positive rate. This surely looks encouraging. We are, however, working with an outpatient clientele only 1/10th of whom are actually schizophrenic. Then, since

$$\Delta p_1 = .20 \quad P = .10$$

$$\Delta p_2 = .05 \quad Q = .90$$

applying the formula we see that

$$\frac{.20}{.05} > \frac{.90}{.10}$$

i.e., the required inequality does not hold, and the routine use of this "improved" index will result in an increase in the proportion of erroneous diagnostic decisions.

In the case of any pair of unimodal distributions, this corresponds to the principle that the optimal cut lies at the intersection of the two distribution envelopes (11, pp. 271-272).

MANIPULATION OF CUTTING LINES FOR DIFFERENT DECISIONS

For any given psychometric device, no one cutting line is maximally efficient for clinical settings in which the base rates of the criterion groups in the population are different. Furthermore, different cutting lines may be necessary for various decisions within the same population. In this section, methods are presented for manipulating the cutting line of any instrument in order to maximize the efficiency of a device in the making of several kinds of decisions. Reference should be made to the scheme presented in Table 5 for understanding of the discussion which follows. This scheme and the methods for manipulating cutting lines are derived from Duncan, Ohlin, Reiss, and Stanton (6).

A study in the prediction of juvenile delinquency by Glueck and Glueck (7) will be used for illustration. Scores on a prediction index for 451 delinquents and 439 nondelinquents (7, p. 261) are listed in Table 6. If the Gluecks' index is to be used in a population with a given juvenile delinquency rate, cutting lines can be established to maximize the efficiency of the index for several de-

TABLE 5

SYMBOLS TO BE USED IN EVALUATING THE EFFICIENCY OF A PSYCHOMETRIC DEVICE
IN CLASSIFICATION OR PREDICTION

| Diagnosis from Test | Actual Diagnosis | | Total Diagnosed from Test |
|--------------------------------|---|---|--|
| | Positive | Negative | |
| Positive | NPp_1 (Number of valid posi- tives) | NQp_2 (Number of false posi- tives) | $NPp_1 + NQp_2$ (Number of test posi- tives) |
| Negative | NPq_1 (Number of false nega- tives) | NQq_2 (Number of valid nega- tives) | $NPq_1 + NQq_2$ (Number of test nega- tives) |
| Total with actual diagnosis | NP (Number of actual posi- tives) | NQ (Number of actual nega- tives) | N (Total number of cases) |

Note.—For simplicity, the term "diagnosis" is used to denote the classification of any kind of pathology, behavior, or event studied, or to denote "outcome" if a test is used for prediction. "Number" means absolute frequency, not rate or probability.

cisions. In the following illustration, a delinquency rate of .20 will be used. From the data in Table 6, optimal cutting lines will be determined for maximizing the proportion of correct predictions, or hits, for all cases (H_T), and for maximizing the proportion of hits (H_P) among those called delinquent (positives) by the index.

In the first three columns of Table 6, " f " denotes the number of delinquents scoring in each class interval, " cf " represents the cumulative

frequency of delinquents scoring above each class interval (e.g., 265 score above 299), and p_1 represents the proportion of the total group of 451 delinquents scoring above each class interval. Columns 4, 5, and 6 present the same kind of data for the 439 nondelinquents.

Maximizing the number of correct predictions or classifications for all cases. The proportion of correct predictions or classifications (H_T) for any given cutting line is given by the formula, $H_T = Pp_1 + Qq_2$. Thus, in

TABLE 6

PREDICTION INDEX SCORES FOR JUVENILE DELINQUENTS AND NONDELINQUENTS AND
OTHER STATISTICS FOR DETERMINING OPTIMAL CUTTING LINES FOR CERTAIN
DECISIONS IN A POPULATION WITH A DELINQUENCY RATE OF .20*

| Prediction Index Score | Delinquents | | | Nondelinquents | | | 1 - p_1 | .2 p_1 | .8 p_1 | .8 q_1 | $Pp_1 + Qq_2$ | $Pp_1 + Qp_1$ | $\frac{Pp_1}{R_P}$ | | | | | | | |
|------------------------------|-------------|------|--------|----------------|------|--------|-----------|----------|----------|----------|---------------|---------------|--------------------|-------|--------|--------|--------|-------|-------|-------|
| | cf | | | cf | | | | | | | | | | | | | | | | |
| | 451 | | | 439 | | | | | | | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | | | | | | | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| | f | cf | p_1 | f | cf | p_1 | | | | | | | | q_1 | Pp_1 | Qp_1 | Qq_1 | H_T | R_P | H_P |
| 400+ | 51 | 51 | .1131 | 1 | 1 | .0023 | .9977 | .0226 | .0018 | .7982 | .821 | .024 | .926 | | | | | | | |
| 350-399 | 73 | 124 | .2749 | 8 | 9 | .0205 | .9795 | .0550 | .0164 | .7836 | .839 | .071 | .770 | | | | | | | |
| 300-349 | 141 | 265 | .5876 | 23 | 32 | .0729 | .9271 | .1175 | .0583 | .7417 | .859 | .176 | .668 | | | | | | | |
| 250-299 | 122 | 387 | .8581 | 70 | 102 | .2323 | .7677 | .1716 | .1858 | .6142 | .786 | .357 | .480 | | | | | | | |
| 200-249 | 40 | 427 | .9468 | 68 | 170 | .3872 | .6128 | .1894 | .3098 | .4902 | .680 | .499 | .379 | | | | | | | |
| 150-199 | 19 | 446 | .9889 | 102 | 272 | .6196 | .3804 | .1978 | .4957 | .3043 | .502 | .694 | .285 | | | | | | | |
| <150 | 5 | 451 | 1.0000 | 167 | 439 | 1.0000 | .0000 | .2000 | .8000 | .0000 | .200 | 1.000 | .200 | | | | | | | |

* Frequencies in columns 1 and 4 are from Glueck and Glueck (7, p. 261).

column 11 of Table 6, labelled H_T , it can be seen that the best cutting line for this decision would be between 299 and 300, for 85.9% of all predictions would be correct if those above the line were predicted to become delinquent and all those below the line nondelinquent. Any other cutting line would result in a smaller proportion of correct predictions, and, in fact, any cutting line set lower than this point would make the index inferior to the use of the base rates, for if all cases were predicted to be nondelinquent, the total proportion of hits would be .80.

Maximizing the number of correct predictions or classifications for positives. The primary use of a prediction device may be for selection of (a) students who will succeed in a training program, (b) applicants who will succeed in a certain job, (c) patients who will benefit from a certain type of therapy, etc. In the present illustration, the index would most likely be used for detection of those who are likely to become delinquents. Thus, the aim might be to maximize the number of hits only within the group predicted by the index to become delinquents (predicted positives = $NPp_1 + NQp_2$). The proportion of correct predictions for this group by the use of different cutting lines is given in column 13, labelled H_P . Thus, if a cutting line is set between 399 and 400, one will be correct over 92 times in 100 if predictions are made only for persons scoring above the cutting line. The formula for determining the efficiency of the test when only positive predictions are made is $H_P = Pp_1 / (Pp_1 + Qp_2)$.

One has to pay a price for achieving a very high level of accuracy with the index. Since the problem is to select potential delinquents so that some sort of therapy can be at-

tempted, the proportion of this selected group in the total sample may be considered as a selection ratio. The selection ratio for positives is $R_P = Pp_1 + Qp_2$, that is, predictions are made only for those above the cutting line. The selection ratio for each possible cutting line is shown in column 12 of Table 6, labelled R_P . It can be seen that to obtain maximum accuracy in selection of delinquents (92.6%), predictions can be made for only 2.4% of the population. For other cutting lines, the accuracy of selection and the corresponding selection ratios are given in Table 6. The worker applying the index must use his own judgment in deciding upon the level of accuracy and the selection ratio desired.

Maximizing the number of correct predictions or classifications for negatives. In some selection problems, the goal is the selection of negatives rather than positives. Then, the proportion of hits among all predicted negative for any given cutting line is $H_N = Qq_2 / (Qq_2 + Pq_1)$, and the selection ratio for negatives is $R_N = Pq_1 + Qq_2$.

In all of the above manipulations of cutting lines, it is essential that there be a large number of cases. Otherwise, the percentages about any given cutting line would be so unstable that very dissimilar results would be obtained on new samples. For most studies in clinical psychology, therefore, it would be necessary to establish cutting lines according to the decisions and methods discussed above, and then to cross validate a specific cutting line on new samples.

The amount of shrinkage to be expected in the cross validation of cutting lines cannot be determined until a thorough mathematical and statistical study of the subject is made. It may be found that when criterion distributions are approxi-

TABLE 7

PERCENTAGE OF DELINQUENTS (D) AND NONDELINQUENTS (ND) IN EACH PREDICTION INDEX SCORE INTERVAL IN A POPULATION IN WHICH THE DELINQUENCY RATE IS .20*

| Prediction Index Score Interval | No. of D | No. of ND | Total of D and ND | % of D in Score Interval | % of ND in Score Interval | % of D and ND in Score Interval |
|---------------------------------|----------|-----------|-------------------|--------------------------|---------------------------|---------------------------------|
| 400+ | 51 | 4 | 55 | 92.7 | 7.3 | 100 |
| 350-399 | 73 | 33 | 106 | 68.9 | 31.1 | 100 |
| 300-349 | 141 | 95 | 236 | 59.7 | 40.3 | 100 |
| 250-299 | 122 | 288 | 410 | 29.8 | 70.2 | 100 |
| 200-249 | 40 | 279 | 319 | 12.5 | 87.5 | 100 |
| 150-199 | 19 | 419 | 438 | 4.3 | 95.7 | 100 |
| <150 | 5 | 686 | 691 | .7 | 99.3 | 100 |
| Total | 451 | 1804 | 2255 | | | |

* Modification of Table XX-2, p. 261, from Glueck and Glueck (7).

mately normal and large, cutting lines should be established in terms of the normal probability table rather than on the basis of the observed p and q values found in the samples. In a later section dealing with the selection ratio we shall see that it is sometimes the best procedure to select all individuals falling above a certain cutting line and to select the others needed to reach the selection ratio by choosing at random below the line; or in other cases to establish several different cuts defining ranges within which one or the opposite decision should be made.

Decisions based on score intervals rather than cutting lines. The Gluecks' data can be used to illustrate another approach to psychometric classification and prediction when scores for large samples are available with a relatively large number of cases in each score interval. In Table 7 are listed frequencies of delinquents and nondelinquents for prediction index score intervals. The frequencies for delinquents are the same as those in Table 6, whereas those for nondelinquents have been corrected for a base rate of .20 by multiplying each

frequency in column 4 of Table 6¹ by

$$4.11 = \frac{(.80) (459)}{(.20) (431)}.$$

Table 7 indicates the proportion of delinquents and nondelinquents among all juveniles who fall within a given score interval when the base rate of delinquency is .20. It can be predicted that of those scoring 400 or more, 92.7% will become delinquent, of those scoring between 350 and 399, 68.9% will be delinquent, etc. Likewise, of those scoring between 200 and 249, it can be predicted that 87.5% will not become delinquent. Since 80% of predictions will be correct without the index if all cases are called nondelinquent, one would not predict nondelinquency with the index in score intervals over 249. Likewise, it would be best not to predict delinquency for individuals

¹ The Gluecks' Tables XX-2, 3, 4, 5, (7, pp. 261-262) and their interpretations therefrom are apt to be misleading because of their exclusive consideration of approximately equal base rates of delinquency and nondelinquency. Reiss (13), in his review of the Gluecks' study, has also discussed their use of an unrepresentative rate of delinquency.

in the intervals under 250 because 20% of predictions will be correct if the base rate is used.

It should be emphasized that there are different ways of quantifying one's clinical errors, and they will, of course, not all give the same evaluation when applied in a given setting. "Per cent valid positives" ($=p_1$) is rarely if ever meaningful without the correlated "per cent false positives" ($=p_2$), and clinicians are accustomed to the idea that we pay for an increase in the first by an increase in the second, whenever the increase is achieved not by an improvement in the test's intrinsic validity but by a shifting of the cutting score. But the two quantities p_1 and p_2 do not define our over-all hit frequency, which depends also upon the base rates P and Q . The three quantities p_1 , p_2 , and P do, however, contain all the information needed to evaluate the test with respect to any given sign or cutting score that yields these values. Although p_1 , p_2 , and P contain the relevant information, other forms of it may be of greater importance. No two of these numbers, for example, answer the obvious question most commonly asked (or vaguely implied) by psychiatrists when an inference is made from a sign, viz., "How sure can you be on the basis of that sign?" The answer to this eminently practical query involves a probability different from any of the above, namely, the *inverse* probability given by Bayes' formula:

$$H_P = \frac{Pp_1}{Pp_1 + Qp_2}.$$

Even a small improvement in the hit frequency to $H'_T = Pp_1 + Qq_2$ over the $H_T = P$ attainable without the test may be adjudged as worth while when the increment ΔH_T is multiplied by the N examined in the course

of one year and is thus seen to involve a dozen lives or a dozen curable schizophrenics. On the other hand, the simple fact that an actual *shrinkage* in total hit rate may occur seems to be unappreciated or tacitly ignored by a good deal of clinical practice. One must keep constantly in mind that numerous diagnostic, prognostic, and dynamic statements can be made about almost all neurotic patients (e.g., "depressed," "inadequate ability to relate," "sexual difficulties") or about very few patients (e.g., "dangerous," "will act out in therapy," "suicidal," "will blow up into a schizophrenia"). A psychologist who uses a test sign that even cross validates at $p_1 = q_2 = 80\%$ to determine whether "depression" is present or absent, working in a clinical population where practically everyone is fairly depressed except a few psychopaths and old-fashioned hysterics, is kidding himself, the psychiatrist, and whoever foots the bill.

"SUCCESSIVE-HURDLES" APPROACH

Tests having low efficiency, or having moderate efficiency but applied to populations having very unbalanced base rates ($P \ll Q$) are sometimes defended by adopting a "crude initial screening" frame of reference, and arguing that certain other procedures (whether tests or not) can be applied to the subset identified by the screener ("successive hurdles"). There is no question that in some circumstances (e.g., military induction, or industrial selection with a large labor market) this is a thoroughly defensible position. However, as a general rule one should examine this type of justification critically, with the preceding considerations in mind. Suppose we have a test which distinguishes brain-tumor from non-brain-tumor pa-

tients with 75% accuracy and no differential bias ($p_1 = q_2 = .75$). Under such circumstances the test hit rate H_T is .75 regardless of the base rate. If we use the test in making our judgments, we are correct in our diagnoses 75 times in 100. But suppose only one patient in 10 actually has a brain tumor, we will drop our over-all "success" from 90% (attainable by diagnosing "No tumor" in all cases) to 75%. We do, however, identify 3 out of 4 of the real brain tumors, and in such a case it seems worth the price. The "price" has two aspects to it: We take time to give the test, and, having given it, we call many "tumorous" who are not. Thus, suppose that in the course of a year we see 1000 patients. Of these, 900 are non-tumor, and we erroneously call 225 of these "tumor." To pick up $(100) (.75) = 75$ of the tumors, *all* 100 of whom would have been called tumor-free using the base rates alone, we are willing to mislabel 3 times this many as tumorous who are actually not. Putting it another way, whenever we say "tumor" on the basis of the test, the chances are 3 to 1 that we are mistaken. When we "rule out" tumor by the test, we are correct 96% of the time, an improvement of only 6% in the confidence attachable to a negative finding over the confidence yielded by the base rates.⁴

Now, picking up the successive-hurdles argument, suppose a major decision (e.g., exploratory surgery) is allowed to rest upon a second test

which is infallible but for practically insuperable reasons of staff, time, etc., cannot be routinely given. We administer Test 2 only to "positives" on (screening) Test 1. By this tactic we eliminate all 225 false positives left by Test 1, and we verify the 75 valid positives screened in by Test 1. The 25 tumors that slipped through as false negatives on Test 1 are, of course, not picked up by Test 2 either, because it is not applied to them. Our total hit frequency is now 97.5%, since the only cases ultimately misclassified out of our 1000 seen are these 25 tumors which escaped through the initial sieve Test 1. We are still running only 7½% above the base rate. We have had to give our short-and-easy test to 1000 individuals and our cumbersome, expensive test to 300 individuals, 225 of whom turn out to be free of tumor. But we have located 75 patients with tumor who would not otherwise have been found.

Such examples suggest that, except in "life-or-death" matters, the successive-screenings argument merely tends to soften the blow of Bayes' Rule in cases where the base rates are very far from symmetry. Also, if Test 2 is not assumed to be infallible but only highly effective, say 90% accurate both ways, results start looking unimpressive again. Our net false positive rate rises from zero to 22 cases miscalled "tumor," and we operate 67 of the actual tumors instead of 75. The total hit frequency drops to 94.5%, only 4½% above that yielded by a blind guessing of the modal class.

THE SELECTION RATIO

Straightforward application of the preceding principles presupposes that the clinical decision maker is free to adopt a policy solely on the basis of maximizing hit frequency. Some-

⁴ Improvements are expressed throughout this article as *absolute* increments in percentage of hits, because: (a) This avoids the complete arbitrariness involved in choosing between original hit rate and miss rate as starting denominator; and (b) for the clinician, the person is the most meaningful unit of gain, rather than a proportion of a proportion (especially when the reference proportion is very small).

times there are external constraints such as staff time, administrative policy, or social obligation which further complicate matters. It may then be impossible to make all decisions in accordance with the base rates, and the task given to the test is that of selecting a subset of cases which are decided in the direction opposite to the base rates but will still contain fewer erroneous decisions than would ever be yielded by opposing the base rates without the test. If 80% of patients referred to a Mental Hygiene Clinic are recoverable with intensive psychotherapy, we would do better to treat everybody than to utilize a test yielding 75% correct predictions. But suppose that available staff time is limited so that we *can* treat only half the referrals. The Bayes-type injunction to "follow the base rates when they are better than the test" becomes pragmatically meaningless, for it directs us to make decisions which we cannot implement. The imposition of an *externally* imposed selection ratio, not determined on the basis of any maximizing or minimizing policy but by nonstatistical considerations, renders the test worth while.

Prior to imposition of any arbitrary selection ratio, the fourfold table for 100 referrals might be as shown in Table 8. If the aim were simply to minimize total errors, we

would predict "good" for each case and be right 80 times in 100. Using the test, we would be right only 75 times in 100. But suppose a selection ratio of .5 is externally imposed. We are then forced to predict "poor" for half the cases, even though this "prediction" is, in any given case, likely to be wrong. (More precisely, we handle this subset *as if* we predicted "poor," by refusing to treat.) So we now select our 50 to-be-treated cases from among those 65 who fall in the "test-good" array, having a frequency of $60/65 = 92.3\%$ hits among those selected. This is better than the 80% we could expect (among those selected) by choosing half the total referrals at random. Of course we pay for this, by making many "false negative" decisions; but these are necessitated, whether we use the test or not, by the fact that the selection ratio was determined without regard for hit maximization but by external considerations. Without the test, our false negative rate q_1 is 50% (i.e., 40 of the 80 "good" cases will be called "poor"); the test reduces the false negative rate to 42.5% ($= 34/80$), since 15 cases from above the cutting line must be selected at random for inclusion in the not-to-be-treated group below the cutting line [i.e., $20 + (60/65)15 = 34$]. Stated in terms of correct decisions, without the test 40 out of 50 selected for therapy will have a good therapeutic outcome; with the test, 46 in 50 will be successes.

Reports of studies in which formulas are developed from psychometrics for the prediction of patients' continuance in psychotherapy have neglected to consider the relationship of the selection ratio to the specific population to which the prediction formula is to be applied. In each study the population has consisted of individuals who were *accepted for*

TABLE 8
ACTUAL AND TEST-PREDICTED
THERAPEUTIC OUTCOME

| Test Prediction | Therapeutic Outcome | | |
|-----------------|---------------------|------|-------|
| | Good | Poor | Total |
| Good | 60 | 5 | 65 |
| Poor | 20 | 15 | 35 |
| Total | 80 | 20 | 100 |

therapy by the usual methods employed at an outpatient clinic, and the prediction formula has been evaluated *only* for such patients. It is implied by these studies that the formula would have the same efficiency if it were used for the *selection* of "continuers" from all those *applying* for therapy. Unless the formula is tested on a random sample of applicants who are allowed to enter therapy without regard to their test scores, its efficiency for selection purposes is unknown. The reported efficiency of the prediction formula in the above studies pertains only to its use in a population of patients who have already been selected for therapy. There is little likelihood that the formula can be used in any practical way for further selection of patients unless the clinic's therapists are carrying a far greater load than they plan to carry in the future.

The use of the term "selection" (as contrasted with "prediction" or "placement") ought not to blind us to the important differences between industrial selection and its clinical analogue. The incidence of false negatives—of potential employees screened out by the test who would actually have made good on the job if hired—is of little concern to management except as it costs money to give tests. Hence the industrial psychologist may choose to express his aim in terms of minimizing the false positives, i.e., of seeing to it that the job success *among those hired* is as large a rate as possible. When we make a clinical decision to treat or not to treat, we are withholding something from people who have a claim upon us in a sense that is much stronger than the "right to work" gives a job applicant any claim upon a particular company. So, even though we speak of a "selection ratio" in clinical work, it must be remem-

bered that those cases *not selected* are patients about whom a certain kind of important negative decision is being made.

For any *given* selection ratio, maximizing total hits is always equivalent to maximizing the hit rate for either type of decision (or minimizing the errors of either, or both, kinds), since cases shifted from one cell of the table have to be exactly compensated for. If m "good" cases that were correctly classified by one decision method are incorrectly classified by another, maintenance of the selection ratio entails that m cases correctly called "poor" are also miscalled "good" by the new method. Hence an externally imposed selection ratio eliminates the often troublesome value questions about the relative seriousness of the two kinds of errors, since they are unavoidably increased or decreased at exactly the same rate.

If the test yields a score or a continuously varying index of some kind, the values of p_1 and p_2 are not fixed, as they may be with "patterns" or "signs." Changes in the selection ratio, R , will then suggest shifting the cutting scores or regions on the basis of the relations obtaining among R , P , and the p_1 , p_2 combinations yielded by various cuts. It is worth special comment that, in the case of continuous distributions, the optimum procedure is *not* always to move the cut until the total area truncated = NR , selecting all above that cut and rejecting all those below. Whether this "obvious" rule is wise or not depends upon the distribution characteristics. We have found it easy to construct pairs of distributions such that the test is "discriminating" throughout, in the sense that the associated cumulative frequencies q_1 and q_2 maintain the same direction of their inequality everywhere in the range

$$\left(\text{i.e., } \frac{1}{N_2} \int_{-\infty}^{x_i} f_2(x) dx \right. \\ \left. > \frac{1}{N_1} \int_{-\infty}^{x_i} f_1(x) dx \text{ for all } x_i \right);$$

yet in which the hit frequency given by a single cut at R is inferior to that given by first selecting with a cut which yields $N_s < NR$, and then picking up the remaining $(NR - N_s)$ cases at random below the cut. Other more complex situations may arise in which different types of decisions should be made in different regions, actually reversing the policy as we move along the test continuum. Such numerical examples as we have constructed utilize continuous, unimodal distributions, and involve differences in variability, skewness, and kurtosis not greater than those which arise fairly often in clinical practice. Of course the utilization of any very complicated pattern of regions requires more stable distribution frequencies than are obtainable from the sample sizes ordinarily available to clinicians.

It is instructive to contemplate some of the moral and administrative issues involved in the practical application of the preceding ideas. It is our impression that a good deal of clinical research is of the "So—what?" variety, not because of defects in experimental design such as inadequate cross validation but because it is hard to see just what are the useful changes in decision making which could reasonably be expected to follow. Suppose, for example, it is shown that "duration of psychotherapy" is 70% predictable from a certain test. Are we prepared to propose that those patients whose test scores fall in a certain range should not receive treatment? If not, then is it of any real advantage therapeutically to "keep in mind" that the pa-

tient has 7 out of 10 chances of staying longer than 15 hours, and 3 out of 10 chances of staying less than that? We are not trying to poke fun at research, since presumably almost any lawful relationship stands a chance of being valuable to our total scientific comprehension some day. But many clinical papers are ostensibly inspired by practical aims, and can be given theoretical interpretation or fitted into any larger framework only with great difficulty if at all. It seems appropriate to urge that such "practical"-oriented investigations should be really *practical*, enabling us to see how our clinical decisions could rationally be modified in the light of the findings. It is doubtful how much of current work could be justified in these terms.

Regardless of whether the test validity is capable of improving on the base rates, there are some prediction problems which have practical import only because of limitations in personnel. What other justification is there for the great emphasis in clinical research on "prognosis," "treatability," or "stayability"? The very formulation of the predictive task as "maximizing the number of hits" already presupposes that we intend *not* to treat some cases; since if we treat all comers, the ascertainment of a bad prognosis score has no practical effect other than to discourage the therapist (and thus hinder therapy?). If intensive psychotherapy could be offered to all veterans who are willing to accept referral to a VA Mental Hygiene Clinic, would it be licit to refuse those who had the poorest outlook? Presumably not. It is interesting to contrast the emphasis on prognosis in clinical psychology with that in, say, cancer surgery, where the treatment of *choice* may still have a very

low probability of "success," but is nevertheless carried out on the basis of that low probability. Nor does this attitude seem unreasonable, since no patient would refuse the best available treatment on the ground that even it was only 10% effective. Suppose a therapist, in the course of earning his living, spends 200 hours a year on nonimprovers by following a decision policy that also results in his unexpected success with one 30-year-old "poor bet." If this client thereby gains $16 \times 365 \times 40 = 233,600$ hours averaging 50% less anxiety during the rest of his natural life, it was presumably worth the price.

These considerations suggest that, with the expansion of professional facilities in the behavior field, the prediction problem will be less like that of industrial *selection* and more like that of *placement*. "To treat or not to treat" or "How treatable" or "How long to treat" would be replaced by "What *kind* of treatment?" But as soon as the problem is formulated in this way, the external selection ratio is usually no longer imposed. Only if we are deciding between such alternatives as classical analysis and, say, 50-hour interpretative therapy would such personnel limitations as can be expected in future years impose an arbitrary *R*. But if the decision is between such alternatives as short-term interpretative therapy, Rogerian therapy, Thorne's directive therapy, hypnotic retraining, and the method of tasks (10, 16, 19), we could "follow the base rates" by treating every patient with the method known to have the highest success frequency among patients "similar" to him. The criteria of similarity (class membership) will presumably be multiple, both phenotypic and genotypic, and will have been chosen be-

cause of their empirically demonstrated prognostic relevance rather than by guesswork, as is current practice. Such an idealized situation also presupposes that the selection and training of psychotherapists will have become socially realistic so that therapeutic personnel skilled in the various methods will be available in some reasonable proportion to the incidence with which each method is the treatment of choice.

How close are we to the upper limit of the predictive validity of personality tests, such as was reached remarkably early in the development of academic aptitude tests? If the now-familiar $\frac{1}{2}$ to $\frac{3}{4}$ proportions of hits against even-split criterion dichotomies are already approaching that upper limit, we may well discover that for many decision problems the search for tests that will significantly better the base rates is a rather unrewarding enterprise. When the criterion is a more circumscribed trait or symptom ("depressed," "affiliative," "sadistic," and the like), the difficulty of improving upon the base rates is combined with the doubtfulness about how valuable it is to have such information with 75% confidence anyhow. But this involves larger issues beyond the scope of the present paper.

AVAILABILITY OF INFORMATION ON BASE RATES

The obvious difficulty we face in practical utilization of the preceding formulas arises from the fact that actual quantitative knowledge of the base rates is usually lacking. But this difficulty must not lead to a dismissal of our considerations as clinically irrelevant. In the case of many clinical decisions, chiefly those involving such phenotypic criteria as overt symptoms, formal diagnosis, subsequent hospitalization, persistence in

therapy, vocational or marital adjustment, and the numerous "surface" personality traits which clinicians try to assess, *the chief reason for our ignorance of the base rates is nothing more subtle than our failure to compute them.* The file data available in most installations having a fairly stable source of clientele would yield values sufficiently accurate to permit minimum and maximum estimates which might be sufficient to decide for or against use of a proposed sign. It is our opinion that this rather mundane taxonomic task is of much greater importance than has been realized, and we hope that the present paper will impel workers to more systematic efforts along these lines.

Even in the case of more subtle, complex, and genotypic inferences, the situation is far from hopeless. Take the case of some such dynamic attribution as "strong latent dependency, which will be anxiety-arousing as therapy proceeds." If this is so difficult to discern *even during intensive therapy* that a therapist's rating on it has too little reliability for use as a criterion, it is hard to see just what is the value of guessing it from psychometrics. If a skilled therapist cannot discriminate the personality characteristic after considerable contact with the patient, it is at least debatable whether the characteristic makes any practical difference. On the other hand, if it can be reliably judged by therapists, the determination of approximate base rates again involves nothing more complex than systematic recording of these judgments and subsequent tabulation. Finally, "clinical experience" and "common sense" must be invoked when there is nothing better to be had. Surely if the q_1/q_2 ratio for a test sign claiming validity for "difficulty in accepting

inner drives" shows from the formula that the base rate must not exceed .65 to justify use of the sign, we can be fairly confident in discarding it for use with *any* psychiatric population! Such a "backward" use of the formula to obtain a maximum useful value of P , in conjunction with the most tolerant common-sense estimates of P from daily experience, will often suffice to answer the question. If one is really in complete ignorance of the limits within which P lies, then obviously no rational judgment as to the probable efficiency of the sign can be made.

ESTIMATION VERSUS SIGNIFICANCE

A further implication of the foregoing thinking is that the exactness of certain small sample statistics, or the relative freedom of certain nonparametric methods from distribution assumptions, has to be stated with care lest it mislead clinicians into an unjustified confidence. When an investigator concludes that a sign, item, cutting score, or pattern has "validity" on the basis of small sample methods, he has rendered a certain very broad null hypothesis unpalatable. To decide, however, whether this "validity" warrants clinicians in using the test is (as every statistician would insist) a further and more complex question. To answer this question, we require more than knowledge that $p_1 \neq p_2$. We need in addition to know, with respect to each decision for which the sign is being proposed, whether the appropriate inequality involving p_1 , p_2 , and P is fulfilled. More than this, since we will usually be extrapolating to a somewhat different clinical population, we need to know whether altered base rates P' and Q' will falsify these inequalities. To do this demands *estimates* of the test param-

ters p_1 and p_2 , the setting up of confidence belts for their difference $p_1 - p_2$ rather than the mere proof of their nonidentity. Finally, if the sign is a cutting score, we will want to consider shifting it so as to maintain optimal hit frequency with new base rates. The effect upon p_1 and p_2 of a contemplated movement of a critical score or band requires a knowledge of distribution form such as only a large sample can give.

As is true in all practical applications of statistical inference, non-mathematical considerations enter into the use of the numerical patterns that exist among P , p_1 , p_2 , and R . But "pragmatic" judgments initially require a separation of the several probabilities involved, some of which may be much more important than others in terms of the human values associated with them. In some settings, over-all hit rate is all that we care about. In others, a redistribution of the hits and misses even without much total improvement may concern us. In still others, the proportions p_1 and p_2 are of primary interest; and, finally, in some instances the confrontation of a certain increment in the absolute frequency (NPp_1) of one group identified will outweigh all other considerations.

Lest our conclusions seem unduly pessimistic, what constructive suggestions can we offer? We have already mentioned the following: (a) Searching for subpopulations with different base rates; (b) successive-hurdles testing; (c) the fact that even a very small percentage of improvement may be worth achieving in certain crucial decisions; (d) the need for systematic collection of base-rate data so that our several equations can be applied. To these we may add two further "constructive" comments. First, test research attention should be largely concen-

trated upon behaviors having base rates nearer a 50-50 split, since it is for these that it is easiest to improve on a base-rate decision policy by use of a test having moderate validity. There are, after all, a large number of clinically important traits which do not occur "almost always" or "very rarely." Test research might be slanted more toward them; the current popularity of Q -sort approaches should facilitate the growth of such an emphasis, by directing attention to items having a reasonable "spread" in the clinical population. Exceptions to such a research policy will arise, in those rare domains where the pragmatic consequences of the alternative decisions justify focusing attention almost wholly on maximizing Pp_1 , with relative neglect of Qp_2 . Secondly, we think the injunction "quit wasting time on noncontributory psychometrics" is really constructive. When the clinical psychologist sees the near futility of predicting rare or near-universal events and traits from test validities incapable of improving upon the base rates, his clinical time is freed for more economically defensible activities, such as research which will improve the parameters p_1 and p_2 ; and for treating patients rather than uttering low-confidence prophecies or truisms about them (in this connection see 12, pp. vii, 7, 127-128). It has not been our intention to be dogmatic about "what is worth finding out, how often." We do suggest that the clinical use of patterns, cutting scores, and signs, or research efforts devoted to the discovery of such, should always be evaluated in the light of the simple algebraic fact discovered in 1763 by Mr. Bayes.

SUMMARY

1. The practical value of a psy-

chometric sign, pattern, or cutting score depends jointly upon its intrinsic validity (in the usual sense of its discriminating power) and the distribution of the criterion variable (base rates) in the clinical population. Almost all contemporary research reporting neglects the base-rate factor and hence makes evaluation of test usefulness difficult or impossible.

2. In some circumstances, notably when the base rates of the criterion classification deviate greatly from a 50 per cent split, use of a test sign having slight or moderate validity will result in an *increase* of erroneous clinical decisions.

3. Even if the test's parameters are precisely known, so that ordinary cross-validation shrinkage is not a problem, application of a sign within a population having these same test parameters but a different base rate may result in a marked change in the proportion of correct decisions. For this reason validation studies should present trustworthy information respecting the criterion distribution in addition to such test parameters as false positive and false negative rates.

4. Establishment of "validity" by exact small sample statistics, since it does not yield accurate information

about the test parameters (a problem of estimation rather than significance), does not permit trustworthy judgments as to test usefulness in a new population with different or unknown base rates.

5. Formulas are presented for determining limits upon relations among (a) the base rates, (b) false negative rate, and (c) false positive rate which must obtain if use of the test sign is to improve clinical decision making.

6. If, however, external constraints (e.g., available staff time) render it administratively unfeasible to decide all cases in accordance with the base rates, a test sign may be worth applying even if following the base rates *would* maximize the total correct decisions, were such a policy possible.

7. Trustworthy information as to the base rates of various patient characteristics can readily be obtained by file research, and test development should (other things being equal) be concentrated on those characteristics having base rates nearer .50 rather than close to .00 or 1.00.

8. The basic rationale is that of Bayes' Theorem concerning the calculation of so-called "inverse probability."

REFERENCES

1. AMERICAN PSYCHOLOGICAL ASSOCIATION, AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AND NATIONAL COUNCIL ON MEASUREMENTS USED IN EDUCATION, JOINT COMMITTEE. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, 51, 201-238.
2. ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology*. (Rev. Ed.) New York: Macmillan, 1949.
3. BROSS, I. D. J. *Design for decision*. New York: Macmillan, 1953.
4. DANIELSON, J. R., & CLARK, J. H. A personality inventory for induction screening. *J. clin. Psychol.*, 1954, 10, 137-143.
5. DORKEN, H., & KRAL, A. The psychological differentiation of organic brain lesions and their localization by means of the Rorschach test. *Amer. J. Psychiat.*, 1952, 108, 764-770.
6. DUNCAN, O. D., OHLIN, L. E., REISS, A. J., & STANTON, H. R. Formal devices for making selection decisions. *Amer. J. Sociol.*, 1953, 58, 573-584.
7. GLUECK, S., & GLUECK, ELEANOR. *Unraveling juvenile delinquency*. Cambridge, Mass.: Harvard Univer. Press, 1950.
8. GOODMAN, L. A. The use and validity of a prediction instrument. I. A reformulation of the use of a prediction instru-

- ment. *Amer. J. Sociol.*, 1953, 58, 503-509.
9. HANVIG, L. J. Some psychological dimensions of low back pain. Unpublished doctor's thesis, Univer. of Minnesota, 1949.
10. HERZBERG, A. *Active psychotherapy*. New York: Grune & Stratton, 1945.
11. HORST, P. (Ed.) The prediction of personal adjustment. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 1-156.
12. MEEHL, P. E. *Clinical versus statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
13. REISS, A. J. Unraveling juvenile delinquency. II. An appraisal of the research methods. *Amer. J. Sociol.*, 1951, 57, 115-120.
14. ROSEN, A. Detection of suicidal patients: an example of some limitations in the prediction of infrequent events. *J. consult. Psychol.*, 1954, 18, 397-403.
15. ROTTER, J. B., RAFFERTY, J. E., & LOT-SOF, A. B. The validity of the Rotter Incomplete Sentences Blank: high school form. *J. consult. Psychol.*, 1954, 18, 105-111.
16. SALTER, A. *Conditioned reflex therapy*. New York: Creative Age Press, 1950.
17. TAULBEE, E. S., & SISSON, B. D. Rorschach pattern analysis in schizophrenia: a cross-validation study. *J. clin. Psychol.*, 1954, 10, 80-82.
18. THIESEN, J. W. A pattern analysis of structural characteristics of the Rorschach test in schizophrenia. *J. consult. Psychol.*, 1952, 16, 365-370.
19. WOLFE, J. Objective psychotherapy of the neuroses. *S. African Med. J.*, 1952, 26, 825-829.

Received for early publication December 24, 1954.

INTRA-INDIVIDUAL RESPONSE VARIABILITY¹

DONALD W. FISKE AND LAURA RICE^{2,3}

University of Chicago

THE PROBLEM

This review of the literature is a foundation stone for a research program on intra-individual variability, the variability of an individual's behavior from one time to another. The program's objectives include the exploration of the phenomena of variability within an individual's behavior, with consequent implications for theory and practice in psychometrics and in personality. The program seeks light on several questions. Can we partial out from the conventional error variance of psychometrics a component of variance over time which is associated with the individual? (Such a component would probably make different proportional contributions to the variance of scores for different individuals.) Are there variability factors analogous to the well-known factors of level scores in mental abilities, interests, and personality? How can the concept of intra-individual variability contribute to problems of the prediction of behavior? What is the signifi-

cance of this concept for the study of personality and for personality theory? Is variability within a class of behavior associated with degree of integration within an area of personality? Are the roots of variability in the neural or physiological functioning of the organism?

The problem of intra-individual variability has not been subjected to systematic conceptualization. Hence it is difficult to set up definite criteria for inclusion in this review and to organize the numerous but highly disparate studies which seem pertinent.

As a means of structuring the problem, we shall first delineate a model instance. *Pure intra-individual variability is defined as the difference between the two responses of an individual at two points in time under the following conditions: (a) the individual is exposed each time to the same stimulus or to objectively indistinguishable stimuli; (b) the total situation in which the responses are made is the same on both occasions.* It is doubtful whether such an abstract case ever exists. Guthrie (89) argues that the exact situation is never reproduced. Several problems await answers. What degree of homogeneity between stimuli and between situations is required before we can assume that psychological equivalence exists? In judging equivalence, what are the boundaries (spatial, temporal, and psychological) of a "total" situation?

In this formulation, the meanings of two words need clarification. We are using "situation" to refer to the total immediate environment of the organism. Strictly speaking, it should

¹ This research was supported in part by the United States Air Force under Contract Number AF 18(600)-601, monitored by the Director of Research, 3300th Research and Development Group (Personnel Research Laboratory), Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

² We are indebted to Dr. Murray Glazer for his helpful comments on a draft of this paper, and to our colleagues: Mr. Fred Zimring helped search the literature; Mr. Jerry Osterweil and Mr. Shib Mitra contributed pertinent criticisms.

³ This review includes publications through 1953.

include the stimulus on which we focus our attention. The situation can be defined as embracing all factors external to the organism, which affect responses.

At this point, "response" is not given a formal definition. When we examine the difference between two responses, we may restrict ourselves to any one of several attributes, such as magnitude, intensity, latency, or quality.

This paradigm involves the further assumption that *the order of the two responses is immaterial*. This requirement implies that the responses show *no systematic trend over time*, due to such processes as learning, fatigue, etc., i.e., that the response is not a function of time. This assumption also implies that the *second response is not affected by either the first response or the first presentation of the stimulus*. We may call this pure case spontaneous variability, as opposed to reactive or adaptive variability.

Finally, there is an assumption which underlies this entire research program: *intra-individual response variability is not random*; it is a lawful phenomenon. The variability of one individual's responses to one stimulus is determined by more or less enduring factors within the individual. Two postulates can be derived from this assumption. (a) Given the same stimulus in the same situation, the difference between an individual's responses at two points in time is related to the difference between his two responses (to that stimulus) at two other points in time. (b) Furthermore, the difference between his responses to one stimulus is related to the difference between responses to at least one other stimulus, objectively distinguishable from the first one. Presumably, the magnitude of this latter relationship is a function of the similarity of the stimuli. This general fundamental

assumption is obviously testable. On the basis of our preliminary studies and of research cited below, we consider that this assumption has been verified.

Data that exactly fit this paradigm are rare. Usually we shall be concerned with variability summed over several stimuli (e.g., number of test responses changed on retest) or with variability in a composite score based on the sum of several responses (e.g., change in total score on an instrument measuring intelligence, performance, interest, etc.).

In the rest of this paper, we shall be primarily concerned with the pure case or approximations to it. Such instances will be called *Type I* or spontaneous variability.

Second in importance to us is *Type II*, the case where all conditions and assumptions for the pure case, Type I, are met with the exception that the sequence of responses shows some pattern or order, other than a monotonic function of time. The simplest example is the alternation of responses. In this class fall instances where the second response is affected by the first response or the first presentation of the stimulus (as in alternation) and also instances where the differences between successive responses are less marked—e.g., where cycles or oscillations are present. We shall not attempt a comprehensive coverage of the literature on Type II variability, especially of the latter kind.

Most instances of Type II variability can be classed as reactive variability: the change in response is determined in part by the organism's reaction to the stimulus it has recently reacted to and/or its reaction to its preceding response.

We do not know yet whether Type I and Type II are actually different in practice. Nevertheless, the concepts underlying these classes are

quite distinct: Type II variability is composed of variability from the same sources as Type I, but includes in addition reactive or periodic variability. Thus all the determinants of Type I variability are present in Type II, but Type II has one extra and major determinant associated with temporal order. For the present, we are making the distinction on an empirical basis: Is there evidence that the responses show some simple ordering? It is possible that the assumption of no systematic trend over time which we make for Type I is unjustified. Cases exist, however, where the order of the responses can be treated as having a negligible effect.

A third class of problems, *Type III*, differs from the pure Type I case in that objectively different stimuli are presented on the two occasions or the background situation is changed. Here, the focus is usually on the appropriateness of the change in response. Is the difference between the subject's two responses too small—does he fail to "adapt" to the change of the stimulus (as in some studies of rigidity)? Is the difference too large—for example, does he over-react to stress?

It is obvious that the Type III case can be taken to include any comparison between responses in two situations which is not included in either Type I or Type II. Hence, our discussion of Type III variability will mention only studies which may help us to understand Type I phenomena. We shall not consider "scatter" or profile variability on different aptitudes. In general, in this paper we are attempting comprehensive coverage of only Type I variability. Our explorations into related topics are solely to clarify our central problem and to obtain leads for attacking it.

As far as we know, there have been no recent reviews of the litera-

ture on Type I variability. Some of the earlier studies are examined by Allport and Vernon (3, pp. 124-128). Solomon (201) discusses research on many topics related to variability, especially work on the avoidance of repetition of response (Type II variability). Glazer (77) examines much of the literature on alternation. An extensive list of papers on behavior in guessing is provided by Senders and Sowards (194). Other surveys of the literature on related topics are mentioned in subsequent sections of this paper.

The sections on these three types of variability are followed by a survey of studies of the correlates of variability. Here are data from which may be gleaned some preliminary notions about the origin and structure of variability.

In this paper, unless otherwise qualified, the term "variability" always refers to intra-individual response variability. Any references to conventional interindividual or group variability (i.e., "individual differences") will be explicitly indicated.

VARIABILITY OF ORGANIC PROCESSES

It is obvious that spontaneous variability (Type I) is produced by factors within the organism. In the pure case, the stimulus and the external situation are unchanged, leaving only the organism as the locus for determinants of variability. Presumably there must be some variation or change in organic processes underlying response variability. Each of these processes may itself be variable in its functioning, or it may be regular over time—we need make no assumption as to the character of this organic variation. However, it would be poor strategy to embark on a study of response variability without evidence that there exists some variabil-

ity in the functioning of organic processes, independent of direct effects of changes in the external environment. This section will cite work which supports such a belief.

Loeb (140, p. 622) points out that interindividual differences and the nonpredictability of individual responses increase as the structure of the organism, especially the nervous system, becomes more complex. Portraying the organism as an active system, Bertalanffy argues that movements (or responses) are influenced by "spontaneous fluctuations in the excitations of nerve centers" (10, p. 167); therefore, from his point of view, reactions are determined by the changing internal situation within the organism, not directly by the stimulus outside. Rashevsky elaborates on these spontaneous fluctuations: "... in general we must expect spontaneous fluctuations of excitation to occur in the central nervous system. Such fluctuations may be due to fluctuations of metabolic activity or to excitation carried to a given region from a number of other regions of the brain, which are randomly excited by the stream of incoming exteroceptive as well as proprioceptive and enteroceptive stimuli" (177, p. 166). (Cf. 130.) Chocholle (32) discusses causes of fluctuation in auditory reaction time which are of central origin, such as the state of the sensory areas.

The spontaneous activity of the nervous system presumably lies behind the extraordinary phenomena reported by Heron, Bexton, and Hebb (105). Subjects kept for a day or more in a condition involving a marked reduction in sensory stimulation reported "visual imagery, dream-like in vividness," akin to hallucination.

Differences in intra-individual variability in biochemistry are emphasized by Jellinek (123). Persky (170)

has compared the average intra-individual variabilities for three biochemical stress indices. A relationship between emotional instability and variability in blood constituents has been reported by Hammett (100) and by Goldstein (82).

In analyzing variation, Crozier and Hoagland (42) distinguish between errors of observation and recording and the real variation found in the responses of living organisms. The latter can be considered to be "unpredictableness" due to the extreme complexity of organic systems. They cite experimental evidence that variability is reduced by an increase in excitation and motor output. We may entertain the tentative hypothesis that, up to some point, as the demands of a stimulus situation increase the mobilization of the organism, the variability of response decreases.

Empirical data on neural and physiological variability are available. Blair and Erlanger (15) found that neural response thresholds and reaction latencies of individual axon fibers vary spontaneously from instant to instant. Herrington (106) measured basal metabolic rate, systolic blood pressure, respiration rate, pulse rate, and rated activity 45 times over 90 days for 11 subjects. With the exception of intercorrelations involving BMR, the median intercorrelation of the standard deviations was as high as the median intercorrelation between means. Furthermore, the ratings on average activity level were related to both the mean levels and the variability of the three physiological measures (again excluding BMR). From an extensive study of basal metabolic rates, Harmon (101) showed that BMR measurements taken under standard conditions indicate considerable variation from day to day.

In his studies of biological intelli-

gence, Halstead (95) has noted that the average deviations of normal subjects on critical fusion frequency show cyclical variation over time. Schmidtke (190) observed that performance on tests of *cff* varied with time of day, the amount of vacillation showing individual differences. McNemar studied variability in critical fusion frequency on different days and under different conditions. She concluded that "individuals do not exhibit the same degree of variability in response errors under all conditions of measurement" (144, p. 21). However, many of the intercorrelations between variability measures from different test conditions on the same day were significant at the .05 level. Variability for the same condition on different days was found to be relatively unstable. This study illustrates a major problem in measuring variability. Were the six observations per trial sufficient to yield a stable value? If not, would it be possible to increase the number of observations and still eliminate any effects of fatigue and lowered motivation?

Fluctuations of minimal sensory stimulations at threshold have been studied for many years. Early work in this area is summarized by Guilford (87), who interpreted them as fluctuations of attention, but Bills (13) objected to this interpretation because the periodicities were different for different sense organs. (Cf. 46, discussed below.)

In this review, we have not attempted systematic coverage of the literature on fluctuations of attention, partly because much of it ignores both inter- and intra-individual differences. For example, Butorin (22) studied variation in speed of addition as a measure of stability of attention.

Brunswick argues that the organism must be flexible because it has such

incomplete knowledge of its environment: "Ambiguity of cues and means relative to the vitally relevant objects and results must find its counterpart in an ambiguity and flexibility of the proximal-peripheral mediating processes in the organism" (21, pp. 257-258). The theme that the capacity to vary responses is essential to individual development and survival is implicit or explicit in many publications. Not only must the individual respond differently to different situations but he must also vary his response to the same situation in order to adapt, i.e., to improve his adjustment to the situation.

PSYCHOMETRIC ASPECTS OF VARIABILITY

Before proceeding with the main body of this paper, the review of studies of individual variability in response, we should consider briefly the various possible measures of variability and some problems associated with this measurement. This discussion is necessary because the topic has not been systematically studied before.

Approaches to the Measurement of Variability

The phenomena of variability are usually viewed negatively. Efforts are usually made to minimize the extent of intra-individual variability. Thus one ordinarily seeks high test-retest reliability (i.e., low variability) and eliminates stimulus items to which inconsistent (varying) responses are made. Our problem can be viewed as the measurement of the unreliability of the individual. Guttman (91) points out that there are three sources of psychometric variation: persons, items, and trials. He formulates an expression for the variance of an individual on an item. In his equation for reliability he includes an error variance term which is the

mean of the error variances for individuals. Thus he does not assume that individuals are equally unreliable, but rather allows for individual differences in unreliability. Guttman (92, 93) has also developed formulas for the reliability of qualitative (categorical) data.

In discussing reliability, Coombs (37) also considers the individual reliability of the same item over time. This formulation is developed in a more generalized form in his *Theory of Psychological Scaling* (38). An earlier paper (35) considers the problem of obtaining a dispersion score for an individual. Both Coombs and Guttman, however, give formal recognition to the concept without solving the practical problems involved in its measurement.

In psychometric theory, the assumption is sometimes made that errors of measurement for two tests or for the same test on two occasions are uncorrelated. This assumption has been questioned by several writers. Thouless (214) discusses and demonstrates the fluctuation of a mental function over time. The variation over time of a person's scores around his true score is noted by Ferguson (68) and by Brown and Thomson (18).

Cattell (24, p. 105) holds that dynamic traits fluctuate even more widely from day to day. (Cf. also 25 and 29.)

An intensive study of components of test unreliability has been made by Fagin (65). He demonstrates that this unreliability is composed of random error variance plus quotidian variance (cf. 228, below) or consistent personal variation. For trait items, more quotidian variance was found in more reliable items, but this relationship was not present in interest items. He concludes that the proportion of quotidian variance can

be estimated and that quotidian scores for people are reliable.

A series of papers by Glaser (78, 79, 80, 81) analyzes changes in the responses on retest. Defining inconsistency as the number of responses changed on retest, he finds that the three intercorrelations of inconsistency scores on three tests (intelligence, interests, personality) are uniformly low although two are statistically significant (78). He indicates that the number of changes between one pair of two trials (out of three) is related to the number between another pair and that the mean difficulty level of items with inconsistent (changed) responses correlates highly with performance (80, 81). (Cf. Yoshioka's finding [231] that rats varied their choice more when the discrimination was more difficult. In a situation where two paths of different length were available, the proportion of choices of the shorter ["correct"] path was a function of the ratio of their lengths, not their absolute lengths.) Glaser shows further that inconsistency scores have no relationship to level scores for a test with an effective range appropriate to the group tested. However, if a test is too easy or too difficult for a group and yields a markedly skewed distribution of scores, a relationship will be found between level score and consistency score. He also reports results consistent with Mosier's finding (157) of high split-half reliability for the difficulty value of the median error. (See also 229, discussed below.)

Cox (39) obtained substantial negative correlations between variability from day to day and initial ability on a motor task. The relation between variability and improvement varied from task to task.

The possibility of a relationship between a variability score and the level score on the same responses

must always be kept in mind. Where the level score contributes to or is a determinant of the variability score, its effects should usually be partialled out. However, it is conceivable that under some conditions, the variability itself may affect the level score.

McReynolds (145) developed a measure of consistency which involved level of difficulty. His subjects were asked whether they could "see" a given concept in an indicated area on a Rorschach card. Ordering the concepts in terms of difficulty, he developed a score based on the relative disorder of the subject's replies, i.e., deviations from saying "yes" to all concepts below a particular level and "no" to all above. This score approximates Coombs' dispersion score of items for one individual (35).

Given a series of repeated measurements for the same individual, several measures of their variability are available and have been used: the standard deviation (42, 136); the average deviation and the range (166). Measures of profile similarity can frequently be used as measures of variability: e.g., *D*, based on squared differences between paired responses, as discussed by Cronbach and Gleser (41). Such a measure may, however, be related to the means and sigmas of the response distributions. Correlation coefficients may also be appropriate measures in some instances.

Noting that variation from day to day occurs even when external conditions are well controlled, Woodrow (228) has suggested the concept of "quotidian variability," which may be used to describe individuals and also to check on the stability of internal conditions during an experiment. He recommends a ratio where the numerator is the standard deviation of the daily means and the denomina-

tor is the average of the standard deviations for each day divided by the square root of the number of trials per day. This formula measures variation from one day to the next in terms of variation within one occasion. However, we must not overlook the individual differences in intra-occasion variability.

Some investigators (64, 72, 116) have utilized measures that take into account the relative position of the responses.

Another method of studying sequences of responses is spectral analysis, which yields a profile of the relative contributions to the total variance (or oscillations of performance) of each of the possible component waves. Abelson (1) demonstrated its use on a perceptual-motor task and compared it to a conventional measure, the variance. The correlation was .08, indicating the essential independence of these two measures of variability derived from the same set of responses.

On the other hand, De Valois (50) found substantial relationships (of the order of .70) among three measures, even when each measure was based on a different set of responses. Using a five-unit maze, his measures were number of different paths used, number of specific choices changed from the choice on last trial, and number of times a third, shorter alternative was used when it was opened up.

The intercorrelation of constancy scores (unchanged responses on retest) has been shown by Dunlap (58) to be in the order of .40 for the several sections of his Preference Blank. The constancy indices had small positive relationships with the level scores for corresponding areas and even less relationship with intelligence.

Many general problems of psychological measurement must be recon-

sidered in developing measures of variability. For example, there is the assumption of additivity of responses which is generally accepted for level scores, in spite of an occasional strong protest. Since we obviously wish to avoid studying the variability of each separate item, we can develop a total change score by counting the number of changed responses (cf. 78, 81, 104, 138). This measure is similar to Zubin's measure of like-mindedness (233).

Such a sum-of-changes score must be distinguished from a score based on change-in-total or level. Thus we can count the number of answers changed on a multiple-choice test or we can compute the change in total score from one trial to the rest (cf. 136 and 206). But a change-in-total score is an index to change in some posited trait, not the tendency to change specific responses.

Another distinction, which was not made in the last section, should be made explicit. Our basic model deals with change in response when the same stimulus is repeated. But there is also the concept of discrepancy between response to homogeneous stimuli, i.e., to stimuli so objectively similar that the same response is expected (cf. Type III variability). Applicable concepts here are Coombs' homogeneity (37) and dispersion score over items (35, 36). Any attempt to measure variability through discrepancies in responses to homogeneous items must be based on the assumption that the degree of homogeneity of two items is constant for all individuals. Otherwise, the discrepancy could be a consequence not of individual variability, but of the heterogeneity of the items for that individual. Thus the study by McReynolds (145), mentioned above, assumes that the concepts had the same order of difficulty or plausibility

for all subjects. Stated in another way, McReynolds did not directly observe intra-individual variability or consistency but rather the extent to which each subject's order of difficulty (as found on one occasion) corresponded with the order of average difficulty for a group.

The Consistency of Variability (Relationships with Time)

Variability involves the difference between two responses. A measure of variability is usually computed from several such differences between paired responses or from the several deviations from the central tendency of a series of responses. The several responses may be made on one occasion, on two occasions, or on several occasions.

Are measures of variability consistent? Does variability within one occasion show internal consistency? Is variability on one occasion related to variability on another? Are variability measures based on comparisons of responses on two or more occasions internally consistent? Do variability measures show systematic trends over time? While definitive answers to all these questions cannot be given at this time, it is appropriate to examine the available evidence.

Some work on variability suggests that variability has fair consistency over time, i.e., variability measures from different occasions are related to each other. Thorndike (213) reported continued variability in the repeated spellings of the same sound in nonsense syllables. In his work on covariation of efficiency of performances on several tasks, Asch (7) noted that an individual's variability remained constant as learning progressed. To increase the stability of his measures of variability, Fryer (73) discarded the first and the last

of ten trials per day. In our analyses of his data, the individual standard deviations for the second, third, and fourth days showed appreciable inter-correlations (.48 to .82) with each other but not with the standard deviations for the first day. Flügel's measure of oscillation (71) was highly consistent from the first half to the second half of 46 daily sessions. (Cf. Lovell [142] discussed below, and De Valois [50] who used an approach analogous to Cronbach's coefficient of stability and equivalence [40, pp. 69-70].)

Scores for fluctuation in attitudes and sentiments between two sessions a day apart were related to scores for fluctuations between sessions a month apart (Cattell, 23): the correlation was .47 for children and .77 for adults. Preliminary studies by the authors suggest that such fluctuation scores are correlated with the tendency to make extreme ratings, but the relationship is not an artifact: unlike those giving many extreme ratings, individuals placing most of their ratings in the center do not change their occasional extreme responses any more than they change their more moderate, center responses.

Cummings (43) measured variability within three-minute periods. For three diverse tests, such scores were highly consistent over ten periods. When the ten periods were on the same day, the average scores for the three tests had only low intercorrelations. However, when the ten periods were on ten different days, the intercorrelations between the average variability scores (for each of the three tests) were sufficiently high to suggest a common variability factor.

The reliability of a variability score may also be a function of the nature of the responses studied. Allport

and Vernon (3) reviewed earlier literature and concluded that variability measures derived from "raw physical scores," such as tapping, are reliable but highly specific, while lower reliability is found for variability measures derived from "more complex tests of intelligence or personality whose initial reliability is not high" (3, p. 128). Also, according to them, Reymert (178) found "that the measure of variability in reaction time failed to correlate with the variabilities of more integrated activities such as reading and counting" (3, p. 132).

A fundamental consideration is the time period between responses. Dudek (56) considers that the variability of individuals often contributes to inconsistency in test scores. He finds that the Spearman-Brown prophecy formula holds for the split-half reliability of a test on a single occasion but may not hold for test-retest reliability. Paulsen (168, 169) also found that split-half reliabilities did not predict test-retest values and reported that intertrial correlations decreased as the number of intervening trials increased. (The latter tendency was also noted by Hertzman [107] and by McNemar [144].) The studies by Dudek and Paulsen used perfectly homogeneous tests of steadiness—the "stimuli" remained the same throughout. Thus, variability over time is related to, but is not identical with, inconsistency of performance on a single occasion.

Variability over time may occur under either of two conditions, defined in terms of rate of organic fluctuation. (a) While the organism may respond consistently within any short test session, it may function differently on different occasions clearly separated in time. (b) On the other hand, the organic processes

determining the response to a given stimulus may be in a continual state of flux, such that two very different responses are made just a few seconds apart. If this condition exists, then the responses within a given occasion are a random sample of the individual's responses to the stimulus and variation over time can be predicted from variation within an occasion. If the first condition holds, variation within an occasion may be unrelated to variation over time.

On the other hand, many studies show that variability itself may be a function of the number of trials: variability may increase or decrease as the series of responses continues. Thorndike (210) reports a slight reduction in variability in drawing hundreds of lines of specified length. With reward or punishment, the average error (variability) in a discrimination task was shown by Hamilton (97) to decrease as the number of trials increased. Lashley (Crozier and Hoagland, 42) found that in archery practice, the standard deviation decreased but the relative variability was constant. For addition problems, Flügel's subjects showed increased absolute oscillation but decreased percentage oscillation over 46 daily sessions (71). Sarvis (188) noted that rhythms in speed of tracing mazes tended to disappear over time. In Hall's study (94), rats showed less variability on later trials in a five-alternative maze. Vacillation decreased with training in the experiment of Mowrer and Jones (159). Variability during conditioning, extinction, and reconditioning was studied by Antonitis (5). He found that variability of response decreased as a function of the number of reinforcements during conditioning, increased rapidly during extinction trials, and decreased during

reconditioning below the level for conditioning trials. Greater variability in output on a continuing task is associated with increased fatigue, according to Bills. This represents a "breakdown in the controlling set" (14, p. 54).

Repeated self-descriptive *Q* sorts have been analyzed by D. M. Taylor (206). In a series of sorts, later ones showed higher intercorrelations than earlier ones. The consistency of self concept over time correlated .33 with the relative positiveness of the initial self-description. Sorts for self were less consistent than sorts for ideals. (Cf. 117, discussed below.)

The *Studies in Expressive Movement*, by Allport and Vernon (3), provide some highly provocative data. They tested 25 subjects on a wide variety of tests in three different sessions. Although they noted that the reliability of level scores was lower between sessions than within sessions, they utilized individual variability scores based on measures from one, two, or three sessions, thus confounding intrasession and intersession variability. Furthermore, while some of the scores were based on variability of response with the stimulus situation essentially unchanged, other variability scores reflected the difference between responses to similar but clearly non-identical stimuli. Also, they utilized several different types of scores, such as average deviation, coefficient of variability, etc. For these reasons, their data are difficult to interpret.

The average intercorrelation among 11 rank orders on variability was .02. They concluded, however, that there was evidence for general variability and specific variability components. The first correlated .26 with their Emphasis factor and the second correlated .38 with Centrifugality.

On the other hand, the data pub-

lished by Hunt (116) show that for both normals and schizophrenics, the rank orders of individuals on variability over 15 trials five days apart maintain some consistency from test to test. (For each group, our analysis of his data yielded a chi square from the analysis of variance of ranks with a p value less than .10.)

Set toward task. Up to this point, we have assumed that the task and the individual's orientation toward it are constant. Some studies, however, suggest that the individual's set toward the task may change as the task proceeds (73, 120), and that differences in set may affect variability.

Abelson (1) reports that the individual variances tended to decrease during the course of the first session but tended to increase during the second session. He interprets the first trend as due to accommodation or learning and the second as due to boredom.

Changes in set occurring later in a series of tests may also affect variability. For rats trained (with food reward) to make brightness discriminations in a maze with a variety of indirect paths, Maier (146) noted tendencies for more errors in later trials on the same day and on later days for the same pattern. He attributed the errors to inattention rather than ignorance. Variability increased when the correct path was left unchanged for several days. Taylor (207) reported a study by Danziger in his laboratory in which the variability of the autokinetic effect remained unchanged over time for volunteer subjects but decreased markedly for paid subjects. In Philpott's work (172), any tendency toward increase in cycle length might well be due to fatigue or boredom (cf. also 127).

The problem of set in relationship to the variability of schizophrenics will be discussed in the section on "Personality types and diagnostic categories."

Summary

The measurement of variability involves a number of problems. Some of them are similar to those present in conventional psychological measurement which emphasizes level scores. Other problems are peculiar to the study of variability. We must guard against implicitly carrying over the assumptions for level scores as we explore this less familiar area.

Because research on variability has been scattered and unsystematic, a number of different measures have been tried, each implying its own concept of variability. We are not yet in a position to decide upon the ideal measure (if one exists).

Are measures of variability within one occasion consistent or stable over time? The answer appears to be a qualified one: under highly constant conditions, variability measures within different occasions may agree well; however, variability measures are readily affected by the set of the subject toward the task, and therefore by changes in such set. On the other hand, measures of variability over occasions may be reasonably stable.

STUDIES OF SPONTANEOUS (APERIODIC) VARIABILITY

In this section we shall consider studies reporting variability which fits closely to our basic paradigm (Type I, spontaneous variability). While such variability has been observed in a wide range of behavioral responses, it has usually been classified as error variance without further interpretation. It is our thesis that this type of variability is a lawful characteristic of an individual in a

situation, and that its investigation will enlarge our understanding of behavior.

The need and the capacity for variation have been distinguished. Many years ago, Tolman emphasized the initial exploratory tendencies in the rat (216) and creative instability, the capacity to break out into new lines of activity (217). Hilgard (108) points out that variability may mean a need or preference for variety and also an ability to vary behavior in a given situation. Both meanings refer to variables underlying the observed phenomena. The need for variety involves a reaction to previous stimuli or responses (cf. Type II). In our approach, we assume that the potentiality for changing a response always exists. Maier (147) also mentions a need for variability and argues that in some instances (e.g., operated rats), the capacity to vary behavior exists but is not used. In a later paper, with Schneirla (149), he again considers the tendency to vary behavior. Mowrer and Jones (159), explored variability as a function of the effort involved in the responses. In Cattell's formulation (25, p. 635), there is a "law of dispersion with excitement and deprivation. Continued stimulation of ergs, with deprivation of the goal, produces increasing variation in the stimuli to which attention is directed and increasing variation of response behavior (as well as introspectively, increased 'excitement')."

In his formal theory, Hull (113, pp. 304-321) gave explicit recognition to the problem of variability by his concept of behavioral oscillation. This oscillation was postulated to be specific to each reaction potential, i.e., the oscillations of different reaction potentials are asynchronous. He also called attention to other

kinds of variability: "The 'constant' numerical values appearing in equations representing primary molar behavioral laws vary from species to species, from individual to individual, and from some physiological states to others in the same individual at different times, all quite apart from the factor of behavioral oscillation (*sOr*)" (114, p. 117). Subsequently, he considered oscillation in relation to conflict situations, behavioral inconsistency in evaluative choices, and alternation tendencies (115). Taylor (207, 208) discussed the concept of behavioral oscillation and argued that it is related to strength of drive.

Statistical learning theory has recognized the problems of stimulus and response variability (Estes, 61; Estes and Burke, 62). Their formulations may lead to a model that can be fitted to the intraorganic determinants of response variability.

Psychophysics would seem at first thought to be an excellent source of data fitting our Type I model of variability, since many of its methods involve repetitions of the same stimulus. However, most of the research in this area is normative—it is concerned with general functions, not differences between individuals. Thus Guilford (88) neglects differences in intra-individual variability. Thurstone (215) points out that the slope of the psychometric function for the constant method indicates degree of sensitivity. Bevan and Dukes (11) show that smaller average errors are obtained in judging the distance of a more valued stimulus object, and Dukes and Bevan (57) found less variability of response in comparing two positively-valued objects than in comparing two negative objects or a positive and a negative object.

Accuracy of performance. Some of the first work (e.g., 210) was on vari-

ability in accuracy of performance. For several studies of repeated testing, Thorndike (211) studied the distribution of the individual's deviation from his own average. Since this study utilized alternate forms of the tests, it does not fully fit the paradigm (cf. 213).

Flügel (71) studied intrasession oscillation and day-to-day variability in continuous addition problems. Absolute oscillation was measured by the sum of changes between successive brief periods and absolute variability was based on deviations from the mean of the five sessions centered around a given day. Using both absolute and relative measures, he found that oscillations and variability were positively associated. He points out that in his data, oscillations usually represented dips from a relatively steady rate, rather than spurts of faster performance.

Hertzman (107) developed a variability score based on deviations from each subject's median on the Thurstone Substitution Test. For a group of highly variable subjects, the size of the intercorrelations between variability scores for different trials was a function of proximity in time. A similar but less marked trend was found in the low variability group.

Variability in hand-arm steadiness within one session was studied by Lovell (142). Three measures were used: average deviation, relative variability (average deviation divided by mean), and sum of successive differences between pairs of trials. For all three, substantial correlations were obtained between scores for two sessions a month apart. The values ranged from .49 to .74, with those for relative variability being the lowest. (Scores were computed for each of three time intervals: 1, 3, and 12 seconds.)

Jarrett (122) reports individual differences in proportions of responses changed on a multiple-choice test.

Output or magnitude of response. The well-known work of Dodge (54) employed a series of measures at different levels of neural integration. Darroch (44) reported daily fluctuations on a perseveration test (where the score is actually based on the difference between performances on two tasks). Johnson (124) found wide individual variation in finger pressure.

Marked individual differences in variability of judgments about a series of colors were reported by Hunt and Flannery (117). Variability increased with the number of colors to be judged and the number of categories used, but decreased as the number of repetitions mounted. David and Rabinowitz (45) developed an instability score based on changes in preference for the Szondi pictures. Our analysis of their data shows substantial correlations between instability scores for the eight "factors" in the test, the average intercorrelation being .42 for nurses and .62 for paranoid schizophrenics. (These are somewhat spurious: because it is a forced-choice test, a change in preference for one picture must produce a change in preference for some other picture.)

Several studies of the reliability of paper-and-pencil inventories have data on variability. Substantial internal consistencies were found by Lentz (138) for number of items changed and for change in total score, these two measures correlating .40. As we might expect, certain items contributed more changes than others, especially the more general and abstract ones. Glaser's criticism (78) that the relationship between level scores and change scores is a function of the range and design of

the test may be applicable to these findings (cf. 163, 173).

Variability has also been noted by students of personality. The reliability of personality measurements is a function of the stability of the individual personality, according to Maller (150). Roshal (185) concludes that successful therapy is associated with gains in behavior variability which contributes to adaptability. Rapaport (176) observes that "variability of reactions" follows from "the multiple determination of psychic events." Saudek (189, p. 62) discusses changes of pressure in handwriting. Spontaneous changes in characteristics of responses within one Rorschach examination are discussed by Beck (9). In a study of changes in responses to a modified Rorschach test, Siipola *et al.* (196) found that almost half the responses were changed more or less. Allen *et al.* (2) report that two-thirds of the responses were changed. Gibby (75) studied the stability of intellectual variables for repeated Rorschach tests. Fluctuations from test to retest on the Bender-Gestalt are reported by Pascal and Suttell (165), who suggest that attitude may be the explanation. It seems reasonable to hypothesize that some variability can be accounted for by change in attitude or set; in fact, it is possible that differences between individuals with respect to variability may be a function of differences in the strength of a set and in the capacity or need to maintain it.

Other work. Several studies cannot be grouped into the categories used above. The variation in response pattern shown by mammals when confronted with an unsolvable task is more a characteristic of the individual than of the species, according to Hamilton (96).

Variability of the instrumental

behavior in guinea pigs was studied by Muenzinger (160), who noted that plasticity remained even after one thousand trials in a puzzle box. Some animals would repeat one movement for a large number of trials and then shift to another one. In a subsequent study (Muenzinger, Koerner, and Irey, 161), animals which were required to make a specific movement in order to escape showed variation in the required movement but also showed a greater number of accessory movements.

Guthrie and Horton (90) observed the behavior of cats in repeated escapes from a puzzle box. They concluded that the cat tends to repeat his previous response. Inspection of their photographic records, however, indicates marked differences between the cats in their degree of consistency.

Rimoldi mentions a relationship between mean reaction time and variation: "a small coefficient of variation and high speed tend to go together, while in slow individuals the variability may be either high or low" (181, p. 298). (Cf. also 136, 146.)

The studies discussed above have demonstrated variability in performance with respect to accuracy, magnitude, and rate, and have illustrated variability in instrumental acts. The evidence justifies the authors' view that significant components of such variability can be isolated and interpreted.

STUDIES OF SYSTEMATIC VARIABILITY

In this section we shall consider investigations using repetitions of the same stimulus in the same situation where the responses show no monotonic trend but where some regularity or order is noted (Type II variability). Thus studies involving improvement and deterioration of the

quality, strength, or rate of response (e.g., due to learning or fatigue) will be omitted, but we shall include data that are viewed as a stationary time series, where the starting point is immaterial. We shall omit references to phenomena showing fixed cycles—such as diurnal fluctuations.

Most examples of systematic variability can be classified as reactive variability—the subject responds not only to the re-presented stimulus but also to the previous presentation or to his previous response to it. In a discussion of difficulties in using repeated measurements, Smith (198) points out that successive responses depend more or less on preceding stimulus-response situations. The problem is made more complex by the fact that the extent to which a response is influenced by preceding responses varies throughout the series, perhaps even from one response to the next. Johnson (according to London, 141) also argues that the state of the organism is altered by the previous response. The non-independence of successive responses in measuring visual threshold was reported and analyzed by Verplanck *et al.* (221). Wertheimer (226) corroborated this result and also found significant interaction variance between days and subjects.

Serial Order

Variability measured by differences between successive responses falls somewhere between the major Type I and Type II classifications. Since the order of the responses is considered, it does not meet a rigorous interpretation of the first category. On the other hand, it does not imply the degree of regularity found in the reactive or periodic forms of systematic variability.

Philip (171) had twelve subjects tap alternately on two plates at

maximum speed until exhausted. In seven subjects he found evidence for periodicity, the periods being about 50 minutes in length. The periodicity was measured by three methods—one of which was the non-randomness of distribution of the most frequently occurring intervals between taps. He distinguished between periodicity (which may be an efficient mode of operation because it provides opportunity for recovery) and variability due to spurts and to the effects of distractions.

Abelson (1) applied spectral analysis to performance on repetitive tasks. This technique assesses the components of the output curve without assuming periodicity. Using an analogy from engineering, he asked whether individuals showed differences in degree and type of "out-of-controlness." (His study does not completely fit our paradigm since his stimuli had some variety, but he assumed with some justification that the analysis could disregard these differences.)

The task was a perceptual-motor one: making 100 jabs at each of five targets. The technique permitted recording the performance with the subject having only minimal knowledge of results.

Abelson obtained repeat measurements on some subjects. As one might expect, the repeat reliability of variability measures for the first task was much lower than for the succeeding four tasks. While the variance for each individual and the value derived from spectral analysis were uncorrelated with each other, each showed stability from test to retest. It may be worth noting that the retested group were volunteers; as a group, they showed on their first test a higher level and a greater scatter of their individual variances

than did the group not volunteering for retest.

The Conceptualization of Systematic Variability

The most prominent figure in the history of work on variability is Raymond Dodge. He was one of the few people to study the general problem. He pointed out that variability could be viewed as accidental deviations from some abstracted "true" measure, but that such an approach with its emphasis upon invariants might overlook significant aspects of behavior (52):

The psychophysical organism is in a perpetual state of flux. . . . Moreover, the neuromuscular consequences of two successive instances of stimulation with physically similar stimuli vary not only according to the momentary conditions of the organism and its psychophysical set, but also according to inner reactions and inhibitions. As is well known, the repetition of identical stimuli may not evoke the same reaction in successive instances (55, p. 5).

Dodge first published his two laws of relative fatigue in 1917. "Within physiological limits, all fatigue decrement in the results of work is relative to the intensity of the stimulus" (51, p. 102). "In any complex of competing tendencies, the relatively greater fatigue of one tendency will tend to eliminate it from the competition in favor of the less fatigued tendencies" (51, p. 105). Recognizing the phenomenon of avoidance of repetition of a response, he suggested the possibility of making use of it to study the nature of inner stimuli. He points out that relative fatigue should be viewed positively, as a conserving mechanism helping to prevent exhaustion.

Dodge (53) collected voluminous data on several levels of neuromuscular processes, for one subject. The records extended over two years and were taken at various times of day,

etc. Unfortunately, Dodge attempted to interpret his data by a series of analogical constructs derived from contemporaneous neurophysiology. Thus he extended the neural concept of refractory phase to reduced responsiveness persisting for minutes and even years.

Solomon (201) notes that Hunter, observing a tendency to alternate (118), assumed it was innate (119). Hull (113) utilized "reactive inhibition" in his system, and Underwood (220) makes this concept the basis of his explanation of alternation. Rothkopf and Zeaman (186) suggest that the alternation tendency has a large "response" component and two small "place" components. Hebb (103, p. 228) has also suggested an interpretation of this phenomenon.

Two papers by Glanzer (76, 77) promulgate the provocative concept of stimulus satiation and give experimental evidence to support it. His basic assumption is that with continued exposure to the same stimuli in the same environment, the organism becomes less active. Every moment the organism perceives a stimulating object, there develops a quantity of stimulus satiation to it. Glanzer maintains that the same principle holds for all multiple-choice and free response situations.

To support his theory, Glanzer provides a telling critique of response theories which are based on the avoidance of the last response. Most of the studies in this area can be interpreted by either theory because it is difficult to determine whether the stimulus or the response is being avoided. There are few crucial experiments other than Glanzer's own work. Although Glanzer does not attempt to do so, it may be possible to adapt his theory to apply to verbal responses. While the nonrepetition of a recent verbal response might

seem to be a case of reactive inhibition, it could be explained as the avoidance of stimulus satiation. If the organism is repeatedly exposed to the same stimulus in the same environment, and if it is forced to respond (i.e., if it cannot become less active), it will introduce variety in its responses. It is obvious that non-repetition of response will reduce boredom.

In this connection, an early study by Robinson and Bills (182) is pertinent. On the basis of introspective reports, they conclude that the rapid repetition of homogeneous responses succeeds best without full attention; subjects were most efficient when they were having "concrete fantasies."

The Nonrepetition of the Preceding Response

A large number of papers deal with the relationship between successive pairs of responses. In a situation with repetitions of the same stimulus and with two or more comparable alternatives, is there a tendency for a response to be different from the preceding one?

Motor responses. Lewin (139) reports an experiment where, with some pressure on them to continue as long as possible, subjects drew repeated moonfaces. While morons either drew them, or broke off to pause or to do something else, normal subjects used secondary activities and other means to keep going.

Most of the work on alternation of motor responses has been done with rats (see reviews in 115 and 201). Yoshioka (230) reported that only a small proportion of his animals distributed their responses evenly between two equally good paths to a goal, and these showed alternation tendencies. MacGillivray and Stone (143) concluded that the tend-

ency to alternate responses is much stronger than the tendency to repeat.

Wingfield and Dennis (227) report 91 per cent alternation for rats given two trials a day but only 68 per cent for six trials a day. Dennis later (48) concluded that alternation effects were due not to a tendency to alternate direction but to a tendency to avoid a specific pathway already taken. However, this latter tendency was not found in mazes with more than two choice points.

Leeper (137) trained rats to choose one path when hungry and the alternative path when thirsty. In considering responses on the second trial on each day, he invoked the idea of a "systematic tendency to variability."

Further light was thrown upon spontaneous alternation by Heathers (102) who found that the percentage of alternation decreased as the time between trials increased from 15 to 120 seconds; when the interval was 15 minutes, alternation disappeared. Weitz and Wakeman (225) reported that alternation decreased to a minimum at intervals of 40 to 50 seconds and then rose with longer intervals. Alternation was noted by Riley and Shapiro (180) for trials 25 seconds apart (but not for trials 5 minutes apart). This tendency declined as the trials continued.

After Solomon's extensive survey of "The Influence of Work on Behavior," he concludes that "work acts to produce negative motivation" (201, p. 35). Solomon (200) was able to increase the percentage of alternation to a maximum of 90 per cent by making rats go up an inclined ramp (according to Zeaman and House, 232). Rothkopf and Zeaman (186) report that alternation was increased by more forced trials and also that it increased as the series of daily trials continued. In

a very recent study using meal worms (85), Grosslight and Ticknor found that alternation is increased by a forced choice and by the summated effect of two preceding turns, but is decreased by a longer distance from the last choice point.

A series of experiments on spontaneous alternation has been reported by Montgomery (152, 153, 154, 155, 156). He believes that spontaneous alternation may be a special case of exploratory tendency. The exploratory tendency is reduced by the exposure to a place (e.g., an arm of a maze), thus leading to an explanation based on place avoidance, not response avoidance. This point of view resembles Glanzer's concept of stimulus satiation (77).

Montgomery concluded that "in simple maze-situations, amount of exploratory behavior decreases as time of exploration increases, and increases proportionally as the area available for exploration increases" (154, p. 584). The same tendency reappeared on each day's trial (155). In both studies, the proportion of alternations in sequences of locomotor responses was above chance.

In another study (153), the percentage of alternation declined as the interval between trials was lengthened. Within each block of ten trials, the alternation rate was constant from the first half to the second. The effort required to press the bar did not affect alternation. The discrepancy between this finding and Solomon's (see above) may be due to any of several factors, such as the kind of work involved.

One group of papers has emphasized the absence of variability. Hamilton and Ellis (98) concluded that normal rats showed persistency in seeking a goal, but variability in their behavioral activities, whereas operated rats showed behavioral con-

stancy—a relatively fixed sequence of behavioral acts functioning as a unit. Hamilton and Krechevsky (99) demonstrated that a shock administered just before the rat reached the choice point was associated with a reduction in variability and a fixation on one choice or the other. In several experiments, the reduction of variability by shock and by conflict was found by Everall (63) who noted a tendency for the same rats to persevere under the different conditions. Krechevsky published three papers on "Brain-Mechanisms and Variability." In the first (132), he found that normal rats used more different paths to reach a goal and shifted the path used more often than did operated rats. In the latter group the size of the lesion was negatively related to the number of paths used. Since he did not count paths on which rats made errors, part of these findings may be a result of the fact that the operated rats made six times as many errors as the normal rats. In the second study (133), normal rats preferred a longer path of varying shape to an alternative, shorter path of fixed length, but operated rats did not. On the other hand, in an experiment where the varying path was much shorter, no significant difference was found between the two groups: both chose the varying, shorter one only slightly more than half the time (134). Krechevsky concluded that operated rats can show variability of response when there is a large difference in the efficiency of the alternatives. However, these are group averages: the operated rats showed a greater tendency to perseveration, to repeat their last response, especially in the second experiment (cf. 147).

Verbal responses. Most studies of patterning in the response sequences of human subjects have used guesses

or psychophysical judgments. Since the literature in this area has been reviewed recently (194, 201), we shall not attempt a complete coverage here.

The nonrandom patterning of verbal responses was noted and discussed by Thorndike (212) and by Dodge (54). In his intensive study of one subject, Dodge (53) included vocal reaction time to a series of words presented in random order. (In written associations, Telford [209] found less repetition with shorter time intervals between responses.) In a study of speed of naming colors Bills (12) believed he had evidence for blocking. Abelson (1) reanalyzed Bills' data and found no rhythmicity.

In a series of five choices between two alternatives, Goodfellow (83) found a tendency to avoid symmetrical patterning. Skinner (197) argued that Goodfellow's data could best be explained as tendency to alternate which is strengthened if the two preceding responses show repetition, and is weakened when they show alternation. Solomon (202) found that alternation tendencies in guessing were unaffected by the interval between guesses or by the effort required to record their guesses.

In a psychophysical experiment using the method of constant stimuli, a tendency to avoid repeating the preceding judgment (especially the judgment of equal) was noted many years ago by Fernberger (69). A similar tendency was reported by Turner (218), by Arons and Irwin (6), and by Irwin and Preston (121). (Cf. 198, discussed above.)

Day (46, as reported by Abelson, 1) studied patterns in differential threshold responses to auditory stimuli and found long successions of correct discriminations and of failures to detect differences. From his reanalysis, Abelson suggests the inter-

pretation that the probability of detecting a difference is greater when a preceding difference has been heard and less when the preceding change was not heard.

Other Work on Systematic Variability

There are several related areas which will be mentioned but will not be comprehensively reviewed.

Periodicity. The search for periodicity in performance has a long history. As early as 1905, Seashore and Kent (193) reported periodic waves in continuous mental work. Sarvis (188) concluded that rhythms were ascertainable in the time required for continuously tracing mazes blindfolded. However, he did not solve the problem of objective tests for rhythmicity: he held that "prolonged experience" was necessary in making judgments about the presence of rhythms.

Philpott has been concerned with this problem for more than twenty years. In an early monograph (172), he provided a history of older work on curves of output and attempted to demonstrate geometric periodicity. He has recently sought to relate his psychological constants to physical constants. Among those unconvinced by Philpott's arguments and concerned about his failure to utilize statistical tests is Richardson (179) who tested a work curve considered representative by Philpott and concluded that its spikiness might be random.

Oscillation. Spearman (203) provided an extended discussion of oscillations in efficiency. He held that these are manifested in fluctuations of minimal sensory impressions, in fluctuations of mental output, and in rivalry such as reversible perspectives. He concluded that there is a general oscillation factor which cannot be explained by g or by persevera-

tion. Tussing (219) found that the rates of fluctuation of four illusions increase with physical fatigue.

Flügel's work on oscillation is discussed above. (Cf. also the section on Serial Order.)

Vacillation. Another more or less tangential topic is vacillation in conflict situations. A systematic discussion is provided by Miller (151). His paradigm is based on the relationship between distance and strength of response tendency. When an approach gradient and an avoidance gradient cross, vacillation of response occurs at the intersection. More vacillation is found in the stable equilibrium produced by two avoidance gradients than in the unstable equilibrium of two approach gradients. More complex situations are also considered.

Systematic covariation. The most extensive work on covariation is that of Cattell. He and his colleagues have reported several studies (27, 28, 29) of *P* technique, the correlation of measures on the same individual made on a number of successive occasions. Fluctuation of attitude seems to be associated with emotionality (low maturity). A central consideration in this work is the relative range of fluctuations in different functions and the influence of that range on the obtained relationships.

Holt (110), following a suggestion from Horn (111), correlated self-ratings and Szondi scores for one subject over twelve trials. A large number of self-rating items had high correlations over trials with Szondi factor *m*.

Some covariation tendency for psychometric tasks was noted by Asch (7); the scores of each individual tended to vary together over time, suggesting a general efficiency factor. Our analyses of his data reveal no relationship between the extents of

total variability on three different tasks.

We have intentionally omitted any consideration of mood swings in both normal and pathological subjects.

In summary, there is considerable evidence of the existence of more or less systematic variability. The most common finding is that among equally efficient response alternatives, a given response will differ from the immediately preceding one, and even from other very recent ones. This is particularly true where the response is a choice of means to the same goal. It also occurs among choices where knowledge of the accuracy of the choice is not available to the subject. There is no general agreement on the explanation for this reactive variability or other systematic variability.

Type II variability can be viewed as a special case of Type I variability. Systematic variability occurs only when the alternative responses have comparable probabilities of occurrence and when the successive presentations of the stimulus are relatively close together in time.

CHANGE OF RESPONSE WITH CHANGE IN STIMULUS OR SITUATION

In this section, we shall consider Type III variability, which we defined earlier as variability in response with variation in the stimulus or in the situation. While this class is obviously very large, some of the research in this area has implications for our central problem. Of course, all variability is of this type if one takes Guthrie's position (89) that the exact situation is never repeated.

Change in Situation Primary

Variability in this category is basically the difference between two

responses to the same stimulus in two different situations.

A great deal of psychological research utilizes the basic design of comparing responses under two different experimental conditions. The interest is generally in the mean change in total score or average performance rather than in absolute change on separate items. Furthermore, the emphasis is usually on the group change, rather than on the distribution of individual change scores. For example, Johnston (125) reports low but generally positive relationships between measures of adjustment and relative gain from test to retest when the retest was administered under stressful conditions. An illustration of a change score based on absolute changes in responses to items is Brownfain's study (20) of the stability of the self concept under varied instructions.

Change in Stimulus Primary

While the concept of variability is closely related to the concept of rigidity, they can be distinguished by two differences in emphasis. In the usual paradigm for studying rigidity, the subject is presented with first one stimulus and then another, objectively different one, with the general situation more or less constant. The focus is on the extent to which the subject fails to change his response, i.e., upon the degree of invariability. In approaching variability, on the other hand, the more typical study keeps both the stimulus and the situation constant and measures the degree of variability, which may be excessively small or large. The second difference is in normative emphasis. Rigidity is conceived to be maladaptive, and hence as the opposite of adaptive, appropriate response tendencies. In studying variability, the adaptiveness of the

change in response is secondary. Change may or may not be considered desirable. On the one hand, fluctuations in quality or accuracy may be viewed as undesirable; on the other hand, the tendency to make use of more than one of several equally efficient alternative responses may be beneficial because it sustains efficiency by preventing boredom. Thus the problem of variability encompasses and is more general than the problem of rigidity.

Krechevsky and Honzik (135) found that individual rats which consistently chose the shorter of two paths in a maze could change, when the paths were interchanged, as rapidly as those showing more variability on the first problem (cf. also 131). In a study of human problem solving, Guetzkow (86) concluded that there are two distinct factors in set: susceptibility to set (tendency to acquire a set readily) and ability to surmount an acquired set. Churchman (33, p. 240) conceives of personality "as the measure (or measures) of typical inefficiency an individual displays in problem-solving," due to his failure to drop an old method or his tendency to change to a new, less efficient method.

Cattell and Tiner (30) have provided a review of the literature on perseveration and a factorial study of structural rigidity. Most of their tests involved capacity to perceive stimuli in new ways, not tendency to respond differently. A further analysis of structural rigidity has been made by Cattell and Winder (31), who distinguish between fluctuations in goals and fluctuations in goal paths. Kleemeier and Dudek (129) sought a general flexibility factor but failed to find it. However, their flexibility tests had little homogeneity.

A major monograph on rigidity is

that by Fisher (70). For him, a measure of rigidity is the number of equivalent alternatives which the subject demonstrate, in some behavioral way, that he can utilize. Once again, the emphasis is upon modes of response which are possible for the subject, not upon variability in actual responses. Thus he used the number of objects liked in a given set of stimuli, the number of possible alternatives accepted by the subject, etc. Several of his suggestions concerning rigidity may have analogues in variability: for example, his discussion of individual consistency of rigidity manifested on tasks of a given difficulty level as related to adjustment, and his distinction between inner rigidity when self-esteem is threatened and peripheral rigidity when not emotionally threatened.

RELATIONSHIPS BETWEEN VARIABILITY AND OTHER VARIABLES

Some Experimental Variables Related to Variability

In this section, we shall review evidence concerning the relationships between variability and each of several kinds of variables. Some studies involving intra-organic conditions will be discussed in this first part.

Brain lesions appear to be associated with reduced variability. Halstead (95) noted that the average deviation of trials on critical fusion frequency trials is lower for frontal lobectomy cases. Reference has been made earlier to several pertinent studies (98, 130, 132, 133, 134).

The effect of shock on variability has been studied by several people (63, 99, 162, 187). Fairlie (66) discovered that shock at the choice point produced more fixation in rats entering the correct path than in rats entering an incorrect one. De Valois (50) offers the interpretation

that fixation occurs when the rat cannot avoid the punishing shock. After discussing the studies (126, 128, 164) that show the fixation persisting after the shock is discontinued, he emphasizes Farber's finding (67) that the persisting fixation is caused by anxiety: when the anxiety state is removed, the fixation is ended.

De Valois' own experiment (50) is a major contribution to this area. He showed that in rats, more intense motivation is associated with lower variability, for both an approach motive (thirst) and an avoidance motive (shock). Increasing the motivation decreased the variability, and decreasing the motivation increased the variability. The latter effect was demonstrated when the original motivation was moderately high but not when it was very strong. De Valois does not accept Maier's position, which includes the principle (148, p. 159) that a problem situation produces stereotyped behavior in a frustrated individual but variable behavior in a motivated one. De Valois' results agree with Elliott's finding (60) that increased hunger lowered variability but that the rats remained at a low level of variability subsequently when they were less hungry.

In human subjects, increase in motivation may increase variability. Deese and Lazarus (47) obtained greater variability of performance on a Rotary Pursuit Test by making the task more important and by inducing failure stress. These factors also increased the interindividual differences in variability. A measure of the variability of reaction time for binocular fusion under conditions involving emotional stress may be useful in selecting emotionally stable people (Brown University, 19). Using only 14 subjects, Baker and

Harris (8) found Rorschach correlates with an increase in variability of speech intensity under stress (threat of shock). It should be noted that these stresses used with human subjects are disruptive in part because no positive adaptation is possible.

The picture at lower levels of motivation is less clear. J. G. Taylor (207) suggested that strength of drive is related to amount of spontaneous activity, one aspect of which is behavioral oscillation. Solomon (201) reported an increase in alternation with increase in effort required. On the other hand, with sets of two equal paths, Mowrer (158, Ch. 6) working with Orbison found that the longer the route, the less the vacillation. This finding may be a function of increased time between trials and/or reduced amounts of stimulus satiation effects (cf. Glanzer, 77); the increase in energy expenditure may be unimportant here.

In an experiment by Goodman, Moyer, and Bunch (84), rats were exposed to electric shock, air blast, or food deprivation before being dropped into water. These various conditions did not produce differences in variability of alley chosen to get out of the water, perhaps because the conditions were not sufficiently stressful.

Several studies have noted changes in variability, both within one occasion and over occasions (1, 94, 159, 207).

Personality as Related to Variability

Personality traits. Hall (94) found that rats showing more variability among five alleys of equal length also showed greater emotionality (as measured by defecation in the apparatus) and took somewhat more time per run. Emotionality might be construed as an index of unreduced tension.

Frenkel-Brunswik (72) reported some provocative relationships between trait ratings and fluctuations in ratings made semiannually.

Cattell (23) confirmed his prediction that degree of change in sentiments and attitudes would be negatively related to *w*, the general character factor, as rated by peers. He explained the lower correlations of his measure of fluctuation with extroversion, emotionality, and mood swings in terms of the relationships between these variables and *w*. Fairly consistent results were obtained for children and for adults, using intervals of either one day or one month between the administrations of the inventories. The *w* factor has its highest correlations with small change in "deeper sentiments." Similarly, Cummings (43) found that variability on self-ratings correlated negatively with persistence, *w*, and introversion. Subjects high on this kind of variability were rated by others as original, imaginative, and talkative; low subjects were regarded as conventional, thorough, and pugnacious. Walton (222) demonstrated a relationship between steadiness of character and low oscillation on motor and cognitive tests.

From a study of weekly retests on the MMPI, Layton (136) concluded that variation is a function of the individual, not of his score relative to the group. Variation on single scales had no consistent relationship to mean score although some trends were in a plausible direction.

Rosenzweig and Mirmow (184) found that degree of socialization was associated with trends on the P-F test, a trend being a shift from extrapunitive to impunitiveness during the testing. (This type of change is presumably due to the subject's reaction to his previous responses.)

Two papers (163, 173) correlated changes in questionnaire responses with test scores on adjustment. Glaser, however, has done a more thorough study (78) demonstrating that the correlations between level scores and consistency are zero for tests with ranges appropriate for the group tested. This criticism does not apply to Weber's study (224) in which variability of speed on psychometric tests was related to emotionality and submissiveness (as measured by questionnaires.)

Several factor analyses including variability measures have been made. Brogden (17) found nonvariability among 30 trials on addition had an appreciable loading on a factor involving ability to work steadily and to resist distraction. In his synthesis of factor studies using objective personality tests, Cattell (26) notes several factors on which variability measures have loadings. Connor (34), using measures of both daily variation and immediate variation, found little relationship between temperament and variability measures from which ability had been partialled out.

Personality integration. In several sources, we find the suggestion that personality integration is negatively related to variability (cf. 23 and 125, discussed above; also 109, 204).

A provocative paper by Smith and Klein (199) indicates a relationship between variability of speed on the Stroop color-name test and a type of adaptability which is accompanied by lability of control.

In their review of the Szondi test, Borstelmann and Klopfer (16) evaluate the proposition that the picture categories in which changes in preference patterns occur represent the more unstable areas of personality, whereas categories representative of stable, basic need systems will show

little change over time in selection pattern (cf. 49). Noting that David and Rabinowitz (45) found greater changes in choices for schizophrenics than for student nurses, Borstelmann and Klopfer conclude that variability in Szondi Test behavior occurs in the records of normal subjects and, to a greater extent, in the response of pathological subjects. For both groups, the "variability seems to be pervasive and not differential among test categories" (16, p. 124).

Personality types and diagnostic categories. Pauli utilized a variety of measures in his analysis of the curve of performance for continuous additions and subtractions (166). In a second paper (167), he concludes that to assess character and aptitude, the total number of items tried and the total number correct are sufficient; the range and average deviation are unnecessary. Susukita (205), using Pauli's methods, found he could distinguish between two Japanese character types: the inner-integrative or rigid had lower variation in performance than the outer-integrative or labile. Variation was not related to age. Andō (4), however, concluded that variability on psychophysical tasks could not be used for character diagnosis. Using instructions to sustain pressure on a dynamometer at a constant level, Eden (59) found that the graph of pressure varied for the different Jaensch types. In general, the J types had smooth curves while those for the S types were irregular.

The variability of performance of normal and neuropsychiatric groups on various tasks has been compared by Eysenck and his colleagues (64) and by Roseman (183).

The question whether psychotics, especially schizophrenics, are more variable than normals has been examined in several papers. The earliest

study was made by Gatewood (74) who found that dementia praecox patients were more irregular than normals. He suggested that the difference was due to the patients' poor attention and their lack of "thought control." Seeking to determine whether schizophrenics were more variable because of defects in "psychological government," Hunt (116) obtained self-reports from his subjects. He found schizophrenics were more variable and that, within that group, output was significantly related to type of preparatory set taken by the subject. The greater variability in output was associated with greater variability in set toward the task. Two other possible governing factors had little or no independent effect.

A similar study was carried out by Huston, Shakow, and Riggs (120). Their schizophrenics had higher means and intra-individual variability in reaction time than their control group. Since the schizophrenics showed lower means and variability in their second and third testing sessions than in their first, the authors suggest that schizophrenics may have a slower rate of adaption. Cooperation and variability were negatively related, but even the more cooperative patients were more variable than the controls. A second experiment tended to confirm the hypothesis that schizophrenic patients do not attain as high a level of preparation or set, show more variability in height of preparation, and do not maintain their best level of preparation as consistently as do normal subjects. An earlier study by Shakow and Huston (195) had shown greater variability for both schizophrenics and manic-depressives than for normals on speed of alternate tapping.

Using tests similar to those of Allport and Vernon (3), Wulfeck (229)

measured intra-individual consistency of performance by test-retest correlations between the second and third sessions. The average r for several tests was .81 for manic-depressives, .80 for normals, .75 for psychoneurotics, and .71 for schizophrenics. Schofield (191, 192) has examined the changes in MMPI responses following different therapies for normals, neurotics, and psychotics. Differences in the intra-individual variability of schizophrenics and controls with respect to physiological functioning are discussed by Hoskins and Jellinek (112).

There have been a number of papers on consistency of intelligence test performance (e.g., 104, 174, 175, 223).

Variability or Variabilities?

With almost no exceptions, each of the studies reviewed in this paper has examined variability on only a single measure. We have no definitive evidence yet on the generality of variability. What is the factor structure of variability scores? Is there one general factor? Are there many common factors?

There is reason to believe that variability will turn out to be a function of the test stimuli and of the general situation. If so, the problem of the correlations between personality and variability becomes more complex. We may have to determine which personality traits are associated with each variability factor or measure.

SUMMARY AND CONCLUSIONS

Intra-individual response variability refers to the change in an individual's response from one time to the next. Variability under three broad classes of conditions has been considered:

Type I. Spontaneous or aperiodic

variability. The individual is presented with the same stimulus in the same situation at two points in time. It is assumed that the initial presentation and the initial response do not affect the second response, i.e., that the order of the responses is immaterial.

Type II. Systematic variability. Although the stimulus and the situation are again unchanged, the second response is influenced by the first presentation, by the first response, or by both. A primary example is reactive variability such as response alternation. Explicitly excluded from this type are (a) change showing a monotonic relationship with time (due to learning, fatigue, etc.) and (b) such periodic and cyclical phenomena as diurnal variation.

Type III. Change in response associated with change in stimulus or in situation. This comprehensive class includes all conditions not mentioned above. Only a few problems in this class, such as rigidity, have pertinence here.

Since the pure case, Type I, excludes all external determinants of change, variability must come from within the organism. We have, therefore, pointed out that fluctuations of intraorganic processes or states have been demonstrated or at least accepted by many writers.

Variability has been measured in many different ways. Furthermore, different variability scores may be based on different time intervals between the responses being compared. Measures of variability from one occasion have reasonable intercorrelations with each other and can have high stability over time, especially over short intervals. In general, we may expect the correlation between two of these measures to be a function of both the time interval

between them and the relative homogeneity of the two tasks from which the scores were obtained. Such a score is itself much more subject to change over time than are conventional level scores. Variability measures appear to be related to set toward a task and to degree of adaptation to the stimulus situation.

Relatively little work has been done on Type I (spontaneous) variability, perhaps because it is rarely seen in its pure form and because it is difficult to elicit in experimental situations. Nevertheless, it has been recognized in almost all kinds of behavior.

Reactive variability (Type II) has been studied more thoroughly. Both motor responses in rats and verbal responses in humans show a tendency toward the nonrepetition of the preceding or very recent responses. It appears likely that the organism does not seek to avoid making the previous response but rather seeks to respond in such a way as to vary the total pattern of stimulation reaching it, including the stimulus produced directly or indirectly by its own response.

In examining some aspects of Type III variability, it was suggested that the concept of rigidity refers to variability which is restricted or reduced to a level considered to be maladaptive.

Variability is probably decreased by shock and by very strong motivation. Motivation at lower levels is less clearly or more complexly related to variation in response, e.g., experimental stress may increase variability.

The personality correlates of variability, if any, remain to be established definitively. It is likely that variability is negatively related to persistence and "character." Simi-

larly, we cannot state with assurance the nature of the relationships with neurosis and psychosis.

This paper has reviewed many studies with more or less relevance

to variability. While sections of the general area have been investigated, few sections have been systematically attacked. The phenomena require further intensive studies.

BIBLIOGRAPHY

- ABELSON, R. P. Spectral analysis and the study of individual differences in the performance of routine, repetitive tasks. Princeton: Educational Testing Service, 1953.
- ALLEN, R. M., MANNE, S. H., & STIFF, MARGARET. The influence of color on the consistency of responses in the Rorschach test. *J. clin. Psychol.*, 1952, **8**, 97-98.
- ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movement*. New York: Macmillan, 1933.
- ANDŌ, M. Abweichung und Variation (Deviation and variance.) *Tohoku Psychol. Folia*, 1942, **9**, 223-232.
- ANTONITIS, J. J. Response variability in the white rat during conditioning, extinction, and reconditioning. *J. exp. Psychol.*, 1951, **42**, 273-281.
- ARONS, L., & IRWIN, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, **15**, 733-756.
- ASCH, S. E. An experimental study of variability in learning. *Arch. Psychol.*, N. Y., 1932, **22**, No. 143.
- BAKER, L. M., & HARRIS, J. S. The validation of the Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, **5**, 161-164.
- BECK, S. J. *Rorschach's Test*. Vol. 2. *A variety of personality pictures*. New York: Grune & Stratton, 1945.
- BERTALANFFY, L. V. *Problems of life*. New York: Wiley, 1952.
- BEVAN, W., JR., & DUKES, W. F. Value and the Weber Constant in the perception of distance. *Amer. J. Psychol.*, 1951, **64**, 580-584.
- BILLS, A. G. Blocking: a new principle of mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230-245.
- BILLS, A. G. *General experimental psychology*. New York: Longmans, Green, 1934.
- BILLS, A. G. *The psychology of efficiency*. New York: Harper, 1943.
- BLAIR, E. A., & ERLANGER, J. A comparison of the characteristics of axons through their individual electric responses. *Amer. J. Physiol.*, 1933, **106**, 524-564.
- BORSTELMANN, L. J., & KLOPPER, W. G. The Szondi Test: a review and critical evaluation. *Psychol. Bull.*, 1953, **50**, 112-132.
- BROGDEN, H. E. A factor analysis of forty character tests. *Psychol. Monogr.*, 1940, **52**, No. 3 (Whole No. 234).
- BROWN, W., & THOMSON, G. H. *The essentials of mental measurement*. Cambridge: Cambridge Univ. Press, 1940.
- BROWN UNIVERSITY. A method of determining reaction time of binocular fusion under conditions of stress. (1942; Publ. Bd. No. 155789) Washington, D. C.: U. S. Dept. Commerce, 1947.
- BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597-606.
- BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255-273.
- BUTORIN, V. I. Disorders of attention in psychoneurotics. *J. gen. Psychol.*, 1938, **18**, 235-251.
- CATTELL, R. B. Fluctuations of sentiments and attitudes as a measure of character integration. *Amer. J. Psychol.*, 1943, **56**, 195-216.
- CATTELL, R. B. *Description and measurement of personality*. Yonkers-on-Hudson: World Book Co., 1946.
- CATTELL, R. B. *Personality: a systematic theoretical and factual study*. New York: McGraw-Hill, 1950.
- CATTELL, R. B. The principal replicated factors in objective personality tests. *J. abnorm. soc. Psychol.*, 1955, **50**, 291-314.
- CATTELL, R. B., CATTELL, A. K. S., & RHYMER, R. M. P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, 1947, **12**, 267-288.
- CATTELL, R. B., & CROSS, K. P. Comparison of the ergic and self-sentiment

- structures found in dynamic traits by R- and P-techniques. *J. Pers.*, 1952, 21, 250-271.
29. CATTELL, R. B., & LUBORSKY, L. P-technique demonstrated as a new clinical method for determining personality and symptom structure. *J. gen. Psychol.*, 1950, 42, 3-24.
 30. CATTELL, R. B., & TIER, L. G. The varieties of structural rigidity. *J. Pers.*, 1949, 17, 321-341.
 31. CATTELL, R. B., & WINDER, A. E. Structural rigidity in relation to learning theory and clinical psychology. *Psychol. Rev.*, 1952, 59, 23-39.
 32. CHOCHOLLE, R. Quelques remarques sur les variations et la variabilité des temps de réaction auditifs. *J. Psychol. norm. path.*, 1948, 41, 345-358.
 33. CHURCHMAN, C. W. *Theory of experimental inference*. New York: Macmillan, 1948.
 34. CONNOR, D. V. Variability in mental test performance. Paper read at Brisbane Branch, Brit. Psychol. Soc., Brisbane, Australia, March, 1953.
 35. COOMBS, C. H. A rationale for the measurement of traits in individuals. *Psychometrika*, 1948, 13, 59-68.
 36. COOMBS, C. H. The measurement of psychological traits, a research program. In Wilma T. Donahue, et al. (Eds.), *The measurement of student adjustment and achievement*. Ann Arbor: Univ. of Michigan Press, 1949. Pp. 227-238.
 37. COOMBS, C. H. The concepts of reliability and homogeneity. *Educ. psychol. Measmt.*, 1950, 10, 43-56.
 38. COOMBS, C. H. A theory of psychological scaling. *Univ. Mich. Engng. Res. Inst. Bull.*, 1951, No. 34.
 39. COX, J. W. *Manual skill*. Cambridge: Cambridge Univ. Press, 1936.
 40. CRONBACH, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
 41. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, 50, 456-473.
 42. CROZIER, W. J., & HOAGLAND, H. The study of living organisms. In C. Murchison (Ed.), *Handbook of general experimental psychology*. Worcester, Mass.: Clark Univ. Press, 1934. Pp. 39-49.
 43. CUMMINGS, JEAN D. Variability of judgment and steadiness of character. *Brit. J. Psychol.*, 1939, 29, 345-370.
 44. DARROCH, J. An investigation into the degree of variation in the score of a motor preservation test. *Brit. J. Psychol.*, 1938, 28, 248-262.
 45. DAVID, H. P., & RABINOWITZ, W. The development of a Szondi Instability Score. *J. consult. Psychol.*, 1951, 15, 334-336.
 46. DAY, W. Stimulus interval as a determiner of serial patterns in threshold responses. Unpublished master's thesis, Univ. of Virginia, 1951.
 47. DEESE, J., & LAZARUS, R. S. The effects of psychological stress upon perceptual-motor performance. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52-19.
 48. DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, 28, 305-312.
 49. DERI, SUSAN. *Introduction to the Szondi Test*. New York: Grune & Stratton, 1949.
 50. DE VALOIS, R. L. The relation of different levels and kinds of motivation to variability of behavior. Unpublished doctor's dissertation, Univ. of Michigan, 1952.
 51. DODGE, R. The laws of relative fatigue. *Psychol. Rev.*, 1917, 24, 89-113.
 52. DODGE, R. Problems of human variability. *Science*, 1924, 59, 263-270.
 53. DODGE, R. *Elementary conditions of human variability*. New York: Columbia Univ. Press, 1927.
 54. DODGE, R. Note on Professor Thorndike's experiment. *Psychol. Rev.*, 1927, 34, 237-240.
 55. DODGE, R. *Conditions and consequences of human variability*. New Haven: Yale Univ. Press, 1931.
 56. DUDEK, F. J. Concerning "reliability" of tests. *Educ. psychol. Measmt.*, 1952, 12, 293-299.
 57. DUKES, W. F., & BEVAN, W., JR. Accentuation and response variability in the perception of personally relevant objects. *J. Pers.*, 1952, 20, 457-465.
 58. DUNLAP, J. W. Relationships between constancy of expressed preferences and certain other factors. *J. educ. Psychol.*, 1936, 27, 521-526.
 59. EDEN, H. Analyse einer einfachen ausseren Willenshandlung unter dem Gesichtspunkt der Integrationstypologie. *Z. Psychol.*, 1937, 141, 198-240.
 60. ELLIOTT, M. H. The effect of hunger on variability of performance. *Amer. J. Psychol.*, 1934, 46, 107-112.
 61. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94-107.

62. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276-286.
63. EVERALL, E. E. Perseveration in the rat. *J. comp. Psychol.*, 1935, **19**, 343-369.
64. EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1947.
65. FAGIN, H. An investigation of the nature and magnitude of the components of retest unreliability. Unpublished doctor's dissertation, New York University, 1950.
66. FAIRLIE, C. W. The effect of shock at the 'moment of choice' on the formation of a visual discrimination habit. *J. exp. Psychol.*, 1937, **21**, 662-669.
67. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111-131.
68. FERGUSON, G. A. *The reliability of mental tests*. London: Univer. of London Press, 1941.
69. FERNBERGER, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, **3**, 126-150.
70. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
71. FLÜGEL, J. C. Practice, fatigue, and oscillation. *Brit. J. Psychol. Monogr. Suppl.*, 1928, **4**, No. 13.
72. FRENKEL-BRUNSWIK, ELSE. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, **26**, 121-265.
73. FRYER, D. Variability in automatic mental performance with uniform intent. *J. appl. Psychol.*, 1937, **21**, 528-545.
74. GATEWOOD, L. C. An experimental study of dementia praecox. *Psychol. Monogr.*, 1909, **11**, No. 2 (Whole No. 45).
75. GIBBY, R. G. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *J. proj. Tech.*, 1951, **15**, 3-25.
76. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387-393.
77. GLANZER, M. Stimulus satiation: a construct to explain spontaneous alternation, variability, and exploratory behavior. *Psychol. Rev.*, 1953, **60**, 257-268.
78. GLASER, R. A methodological analysis of the inconsistency of response to test items. *Educ. psychol. Measmt.*, 1949, **9**, 727-739.
79. GLASER, R. Multiple operation measurement. *Psychol. Rev.*, 1950, **57**, 241-252.
80. GLASER, R. The application of the concepts of multiple operation measurement to the response patterns on psychological tests. *Educ. psychol. Measmt.*, 1951, **11**, 372-382.
81. GLASER, R. The reliability of inconsistency. *Educ. psychol. Measmt.*, 1952, **12**, 60-64.
82. GOLDSTEIN, H. The biochemical variability of the individual in relation to personality and intelligence. *J. exp. Psychol.*, 1935, **18**, 348-371.
83. GOODFELLOW, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601-632.
84. GOODMAN, R. W., MOYER, K. E., & BUNCH, M. E. Variability and behavior constancy in white rats. *J. comp. physiol. Psychol.*, 1952, **45**, 460-467.
85. GROSSLIGHT, J., & TICKNOR, W. Variability and reactive inhibition in the meal worm as a function of determined turning sequences. *J. comp. physiol. Psychol.*, 1953, **46**, 35-38.
86. GURTZKOW, H. An analysis of the operation of set in problem-solving behavior. *J. gen. Psychol.*, 1951, **45**, 219-244.
87. GUILFORD, J. P. Fluctuations of attention with weak visual stimuli. *Amer. J. Psychol.*, 1927, **38**, 534-583.
88. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
89. GUTHRIE, E. R. Personality in terms of associative learning. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 49-68.
90. GUTHRIE, E. R., & HORTON, G. P. *Cats in a puzzle box*. New York: Rinehart, 1946.
91. GUTTMAN, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255-282.
92. GUTTMAN, L. The test-retest reliability of qualitative data. *Psychometrika*, 1946, **11**, 81-95.
93. GUTTMAN, L. Problems of reliability. In S. A. Stouffer, et al. (Eds.), *Measurement and prediction*. (Vol. IV of *Studies in social psychology in World War II*.) Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 302-311.
94. HALL, C. S. Emotional behavior in the rat: IV. The relationship between emo-

- tionality and stereotyping of behavior. *J. comp. Psychol.*, 1937, **24**, 369-375.
95. HALSTEAD, W. C. *Brain and intelligence*. Chicago: Univ. of Chicago Press, 1947.
 96. HAMILTON, G. V. A study of perseverance reactions in primates and rodents. *Behav. Monogr.*, 1916, **3**, No. 2.
 97. HAMILTON, H. C. The effect of incentives on accuracy of discrimination measured on the Galton bar. *Arch. Psychol.*, 1929, **16**, No. 103.
 98. HAMILTON, J. A., & ELLIS, W. D. Persistence and behavior constancy. *J. genet. Psychol.*, 1933, **42**, 140-153.
 99. HAMILTON, J. A., & KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237-253.
 100. HAMMETT, F. S. Observations on the relation between emotional and metabolic stability. *Amer. J. Physiol.*, 1920, **53**, 307-311.
 101. HARMON, F. L. The reliability of metabolism measurements by the closed circuit method. *J. appl. Physiol.*, 1953, **12**, 773-778.
 102. HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359-380.
 103. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
 104. HEIM, A., & WALLACE, J. The effects of repeatedly retesting the group on the same intelligence test. II. High grade mental defectives. *Quart. J. exp. Psychol.*, 1950, **2**, 19-32.
 105. HERON, W., BEXTON, W. H., & HEBB, D. O. Cognitive effects of a decreased variation in the sensory environment. *Amer. Psychologist*, 1953, **8**, 366. (Abstract)
 106. HERRINGTON, L. P. The relation of physiological and social indices of activity level. In Q. McNemar & Maud A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 125-146.
 107. HERTZMAN, M. The influence of the individual's variability on the organization of performance. *J. gen. Psychol.*, 1939, **20**, 3-24.
 108. HILGARD, E. R. *Theories of learning*. New York: Appleton-Century-Cro 1948.
 109. HOLLINGWORTH, H. L. Correlations of achievement within an individual. *J. exp. Psychol.*, 1925, **8**, 190-208.
 110. HOLT, R. R. An approach to the validation of the Szondi Test through a systematic study of unreliability. *J. proj. Tech.*, 1950, **14**, 435-444.
 111. HORN, D. Intra-individual variability in the study of personality. *J. clin. Psychol.*, 1950, **6**, 43-47.
 112. HOSKINS, R. G., & JELLINEK, E. M. The schizophrenic personality with special regard to psychologic and organic concomitants. *Proc. Ass. Res. nerv. ment. Dis.*, 1933, **14**, 211-233.
 113. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
 114. HULL, C. L. *Essentials of behavior*. New Haven: Yale Univ. Press, 1951.
 115. HULL, C. L. *A behavior system*. New Haven: Yale Univ. Press, 1952.
 116. HUNT, J. McV. Psychological government and the high variability of schizophrenic patients. *Amer. J. Psychol.*, 1936, **48**, 64-81.
 117. HUNT, W. A., & FLANNERY, J. Variability in the affective judgment. *Amer. J. Psychol.*, 1938, **51**, 507-513.
 118. HUNTER, W. S. The auditory sensitivity of the white rat. *J. anim. Behav.*, 1914, **4**, 215-222.
 119. HUNTER, W. S. The temporal maze and kinaesthetic sensory processes in the white rat. *Psychobiology*, 1920, **2**, 1-17.
 120. HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39-82.
 121. IRWIN, F. W., & PRESTON, M. G. Avoidance of repetition of judgments across sense modalities. *J. exp. Psychol.*, 1937, **21**, 511-520.
 122. JARRETT, R. F. The extra-chance nature of changes in student's responses to objective test-items. *J. gen. Psychol.*, 1948, **38**, 243-250.
 123. JELLINEK, E. M. The function of biometric methodology in psychiatric research. In *Mental health*, Publication of the AAAS. Lancaster: Science Press, 1939. Pp. 48-59.
 124. JOHNSON, B. Changes in muscular tension in coordinated hand movements. *J. exp. Psychol.*, 1928, **11**, 329-341.
 125. JOHNSTON, J. J. Relationship between emotional adjustment and change scores on speed tests. Unpublished doctor's dissertation, Univ. of Southern California, 1948.
 126. KLEE, J. B. The relation of frustration and motivation to the production of abnormal fixations in the rat. *Psychol. Monogr.*, 1944, **56**, No. 4 (Whole No. 257).
 127. KLEEMEIER, L. B., & KLEEMEIER, R. W. Effects of benzedrine sulphate (am-

- phetamine) on psychomotor performance. *Amer. J. Psychol.*, 1947, **60**, 89-100.
128. KLEEMEIER, R. W. Fixation and regression in the rat. *Psychol. Monogr.*, 1942, **54**, No. 4 (Whole No. 246).
 129. KLEEMEIER, R. W., & DUDEK, F. J. A factorial investigation of flexibility. *Educ. psychol. Measmt*, 1950, **10**, 107-118.
 130. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118-148.
 131. KRECH, D. Dynamic systems as open neurological systems. *Psychol. Rev.*, 1950, **57**, 345-361.
 132. KRECHEVSKY, I. Brain-mechanisms and variability. I. Variability within a means-end-readiness. *J. comp. Psychol.*, 1937, **23**, 121-138.
 133. KRECHEVSKY, I. Brain-mechanisms and variability. II. Variability where no learning is involved. *J. comp. Psychol.*, 1937, **23**, 139-159.
 134. KRECHEVSKY, I. Brain-mechanisms and variability. III. Limitations of the effect of cortical injury upon variability. *J. comp. Psychol.*, 1937, **23**, 351-364.
 135. KRECHEVSKY, I., & HONZIK, C. H. Fixation in the rat. *Univ. Calif. Pub. Psychol.*, 1932, **6**, 13-26.
 136. LAYTON, W. L. The variability of individuals' scores upon successive testings on the Minnesota Multiphasic Personality Inventory. *Amer. Psychologist*, 1952, **7**, 384. (Abstract)
 137. LEEPER, R. The role of motivation in learning: a study of the phenomena of differential motivational control in the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3-40.
 138. LENTZ, T. F., JR. The reliability of opinionnaire technique studied intensively by the retest method. *J. soc. Psychol.*, 1934, **5**, 338-364.
 139. LEWIN, K. *A dynamic theory of personality*. (Trans. by D. K. Adams and K. E. Zener.) New York: McGraw-Hill, 1935.
 140. LOEB, L. *The biological basis of individuality*. Springfield, Ill.: C. C. Thomas, 1945.
 141. LONDON, I. D. Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychol. Rev.*, 1946, **53**, 170-188.
 142. LOVELL, C. A study of personal variation in hand-arm steadiness. *Amer. J. Psychol.*, 1941, **54**, 230-236.
 143. MACGILLIVRAY, M. E., & STONE, C. P. Suggestions toward an explanation of systematic errors made by albino rats in the multiple light discrimination apparatus. *J. genet. Psychol.*, 1930, **38**, 484-489.
 144. MCNEMAR, OLGA W. The ordering of individuals in critical flicker frequency under different measurement conditions. *J. Psychol.*, 1951, **32**, 3-24.
 145. McREYNOLDS, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, **46**, 131-141.
 146. MAIER, N. R. F. Attention and inattention in rats. *J. genet. Psychol.*, 1930, **38**, 288-306.
 147. MAIER, N. R. F. The specific process constituting the learning function. *Psychol. Rev.*, 1939, **46**, 241-252.
 148. MAIER, N. R. F. *Frustration*. New York: McGraw-Hill, 1949.
 149. MAIER, N. R. F., & SCHNEIERLA, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, **49**, 117-134.
 150. MALLER, J. B. Personality tests. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944. Pp. 170-213.
 151. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944. Pp. 431-465.
 152. MONTGOMERY, K. C. An experimental study of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41**, 39-51.
 153. MONTGOMERY, K. C. "Spontaneous alternation" as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42**, 82-93.
 154. MONTGOMERY, K. C. The relation between exploratory behavior and spontaneous alternation in the white rat. *J. comp. physiol. Psychol.*, 1951, **44**, 582-589.
 155. MONTGOMERY, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50-57.
 156. MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287-293.
 157. MOSIER, C. I. Psychophysics and mental test theory. II. The constant process. *Psychol. Rev.*, 1941, **48**, 235-249.
 158. MOWLER, O. H. *Learning theory and per-*

- sonality dynamics. New York: Ronald Press, 1950.
159. MOWRER, O. H., & JONES, HELEN M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369-386.
 160. MUENZINGER, K. F. Plasticity and mechanization of the problem box habit in guinea pigs. *J. comp. Psychol.*, 1928, **8**, 45-70.
 161. MUENZINGER, K. F., KOERNER, L., & IREY, E. Variability of an habitual movement in guinea pigs. *J. comp. Psychol.*, 1929, **9**, 425-436.
 162. MUENZINGER, K. F., & WOOD, A. Motivation in learning. IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J. comp. Psychol.*, 1935, **20**, 95-106.
 163. NEPRASH, J. D. Reliability of questions in the Thurstone Personality Schedule. *J. soc. Psychol.*, 1936, **7**, 239-244.
 164. O'KELLY, L. I. An experimental study of regression. I. Behavioral characteristics of the regressive response. *J. comp. Psychol.*, 1940, **30**, 41-53.
 165. PASCAL, G. R., & TUTTLE, B. J. *The Bender-Gestalt Test*. New York: Grune & Stratton, 1951.
 166. PAULI, R. Beiträge zur Kenntnis der Arbeitskurve. *Arch. ges. Psychol.*, 1936, **97**, 465-532.
 167. PAULI, R. Die Arbeitskurve als ganzheitlicher Prüfungsversuch (als Universaltest). *Arch. ges. Psychol.*, 1938, **100**, 401-423.
 168. PAULSEN, G. B. The reliability and consistency of individual differences in motor control. I. *J. appl. Psychol.*, 1935, **19**, 29-42.
 169. PAULSEN, G. B. The reliability and consistency of individual differences in motor control. II. *J. appl. Psychol.*, 1935, **19**, 166-179.
 170. PERSKY, H. Response to a life stress: evaluation of some biochemical indices. *J. appl. Physiol.*, 1953, **6**, 369-374.
 171. PHILIP, B. R. Studies in high speed continuous work. I. Periodicity. *J. exp. Psychol.*, 1939, **24**, 499-509.
 172. PHILPOTT, S. J. F. Fluctuations in human output. *Brit. J. Psychol. Monogr. Suppl.*, 1933, **6**, No. 17.
 173. PINTNER, R., & FORLANO, G. Four retests of a personality inventory. *J. educ. Psychol.*, 1938, **29**, 93-100.
 174. RABIN, A. I. Fluctuations in the mental level of schizophrenic patients. *Psychiat. Quart.*, 1944, **18**, 78-81.
 175. RABIN, A. I. Test constancy and variation in the mentally ill. *J. gen. Psychol.*, 1944, **31**, 231-239.
 176. RAPAPORT, D. *Emotions and memory*. New York: International Universities Press, 1950.
 177. RASHEVSKY, N. *Advances and applications of mathematical biology*. Chicago: Univ. of Chicago Press, 1940.
 178. REYMERT, M. L. The personal equation in motor capacities. *Scandin. Sci. Rev.*, 1923, **2**, 177-222.
 179. RICHARDSON, L. F. Dr. S. J. F. Philpott's wave theory. *Brit. J. Psychol.*, 1952, **43**, 169-176.
 180. RILEY, D. A., & SHAPIRO, A. N. Alternation behavior as a function of effortfulness of task and distribution of trials. *J. comp. physiol. Psychol.*, 1952, **45**, 468-475.
 181. RIMOLDI, H. J. A. Personal tempo. *J. abnorm. soc. Psychol.*, 1951, **46**, 283-303.
 182. ROBINSON, E. S., & BILLS, A. G. Two factors in the work decrement. *J. exp. Psychol.*, 1926, **9**, 415-443.
 183. ROSEMAN, M. Differential flexibility in normal and neuropsychiatric subjects on a test of closure behavior. Paper read at meeting of South. Soc. for Phil. and Psychol., Roanoke, Va., 1951.
 184. ROSENZWEIG, S., & MIRMOW, E. L. The validation of trends in the children's form of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 306-314.
 185. ROSHAL, J. M. G. Changes in behavior variability with psychotherapy. *Amer. Psychologist*, 1952, **7**, 343. (Abstract)
 186. ROTHKOFF, E. Z., & ZEAMAN, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34**, 235-255.
 187. SANDERS, M. J. An experimental demonstration of regression in the rat. *J. exp. Psychol.*, 1937, **21**, 493-510.
 188. SARVIS, B. C. An experimental study of rhythms. *Psychol. Monogr.*, 1933, **44**, No. 1 (Whole No. 197).
 189. SAUDEK, R. *The psychology of handwriting*. New York: George H. Doran Co., 1926.
 190. SCHMIDTKE, H. Über die Messung der psychischen Ermüdung mit Hilfe des Flimmertests. *Psychol. Forsch.*, 1951, **23**, 409-463.
 191. SCHOFIELD, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol.*

- Monogr.*, 1950, 64, No. 5 (Whole No. 311).
192. SCHOFIELD, W. A further study of the effects of therapies upon MMPI responses. *J. abnorm. soc. Psychol.*, 1953, 48, 67-77.
 193. SEASHORE, C. E., & KENT, G. H. Periodicity and progressive change in continuous mental work. *Psychol. Rev. Monogr. Suppl.*, 1905, 6, 47-101.
 194. SENDERS, V. L., & SOWARDS, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, 65, 358-374.
 195. SHAKOW, D., & HUSTON, P. E. Studies of motor function in schizophrenia. I. Speed of tapping. *J. gen. Psychol.*, 1936, 15, 63-106.
 196. SIIPOLA, ELSA, KUHN, FLORENCE, & TAYLOR, VIVIAN. Measurement of the individual's reaction to color in ink blots. *J. Pers.*, 1950, 19, 153-171.
 197. SKINNER, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, 30, 495-503.
 198. SMITH, B. B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.*, 1951, 42, 250-268.
 199. SMITH, G. J., & KLEIN, G. S. Cognitive controls in serial behavior. *J. Pers.*, 1953, 22, 188-213.
 200. SOLOMON, R. L. The influence of work on behavior. Unpublished doctor's dissertation, Brown Univer., 1947.
 201. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, 45, 1-40.
 202. SOLOMON, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, 39, 322-326.
 203. SPEARMAN, C. *The abilities of man*. New York: Macmillan, 1927.
 204. STALNICK, S. A study of correlation between self-consistency and positive attitudes toward self. Unpublished master's thesis, Univer. of Chicago, 1953.
 205. SUSUKITA, T. Eine Untersuchung über die Struktur des Charakters des japanischen Schulkindes. *Tohoku Psychol. Folia*, 1942, 9, 1-42.
 206. TAYLOR, D. M. Consistency of the self-concept. Unpublished doctor's dissertation, Vanderbilt Univer., 1953.
 207. TAYLOR, J. G. Behavior oscillation and the growth of preference. *Psychol. Rev.*, 1949, 56, 77-87.
 208. TAYLOR, J. G. Reaction latency as a function of reaction potential and behavior oscillation. *Psychol. Rev.*, 1950, 57, 375-389.
 209. TELFORD, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, 14, 1-36.
 210. THORNDIKE, E. L. *The fundamentals of learning*. New York: Teachers College, Columbia Univer., 1916.
 211. THORNDIKE, E. L. The variability of an individual in repetitions of the same task. *J. exp. Psychol.*, 1923, 6, 161-167.
 212. THORNDIKE, E. L. The refractory period in associative processes. *Psychol. Rev.*, 1927, 34, 234-236.
 213. THORNDIKE, E. L. *Human learning*. New York: Appleton-Century-Crofts, 1931.
 214. THOULESS, R. H. Test unreliability and function fluctuation. *Brit. J. Psychol.*, 1936, 26, 325-343.
 215. THURSTONE, L. L. Psychophysical methods. In T. G. Andrews (Ed.), *Methods of psychology*. New York: Wiley, 1948. Pp. 124-157.
 216. TOLMAN, E. C. Purpose and cognition: the determiners of animal learning. *Psychol. Rev.*, 1925, 32, 285-297.
 217. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
 218. TURNER, W. D. Intra-serial effects with lifted weights. *Amer. J. Psychol.*, 1931, 43, 1-25.
 219. TUSSING, L. Perceptual fluctuations of illusions as a possible physical fatigue index. *J. exp. Psychol.*, 1941, 29, 85-88.
 220. UNDERWOOD, B. J. Learning. *Ann. Rev. Psychol.*, 1953, 4, 31-58.
 221. VERPLANCK, W. S., COLLIER, G. H., & COTTON, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, 44, 273-282.
 222. WALTON, R. D. Relations between amplitude of oscillations in short period efficiency and steadiness of character. *Brit. J. Psychol.*, 1936, 27, 181-188.
 223. WEBB, W. B., & DE HAAN, H. Wechsler-Bellevue split-half reliabilities in normals and schizophrenics. *J. consult. Psychol.*, 1951, 15, 68-71.
 224. WEBER, C. O. Function-fluctuation and personality trends of normal subjects. *Amer. J. Psychol.*, 1938, 51, 702-708.
 225. WEITZ, J., & WAKEMAN, M. L. "Spon-

- taneous" alternation and the conditioned response. *J. comp. Psychol.*, 1941, **32**, 551-562.
226. WERTHEIMER, M. An investigation of the "randomness" of threshold measurements. *J. exp. Psychol.*, 1953, **45**, 294-303.
227. WINGFIELD, R. C., & DENNIS, W. The dependence of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135-147.
228. WOODROW, H. Quotidian variability. *Psychol. Rev.*, 1932, **39**, 245-256.
229. WULFECK, W. H. Motor function in the mentally disordered. III. Intra-individual consistency of expressive movement in psychotics, psychoneurotics, and normals. *J. Psychol.*, 1941, **11**, 151-160.
230. YOSHIOKA, J. G. An alternation habit in rats in a simple maze. *J. genet. Psychol.*, 1929, **36**, 257-266.
231. YOSHIOKA, J. G. Weber's law and the discrimination of maze distance in the white rat. *Univer. Calif. Publ. Psychol.*, 1929, **4**, 155-184.
232. ZEAMAN, D., & HOUSE, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177-186.
233. ZUBIN, J. A technique for measuring like-mindedness. *J. abnorm. soc. Psychol.*, 1938, **33**, 508-516.

Received June 10, 1954.

SPECIAL REVIEW

SOME RECENT BOOKS ON COUNSELING AND ADJUSTMENT

EDWARD JOSEPH SHOEN, JR.

Teachers College, Columbia University

- ARBUCKLE, D. S. *Student personnel services in higher education*. New York: McGraw-Hill, 1953. Pp. x+352. \$4.50.
- ARSENIAN, S., & MCKENZIE, F. W. *Counseling in the YMCA*. New York: Association Press, 1954. Pp. 126. \$2.00.
- CRAIG, R. C. *The transfer value of guided learning*. New York: Bureau of Publications, Teachers College, 1953. Pp. viii+85. \$2.75.
- DRIVER, HELEN IRENE. *Multiple counseling*. Madison, Wis.: Monona Publications, 1954. Pp. 280. \$5.00.
- GRIFFITHS, W. *Behavior difficulties of children as perceived and judged by parents, teachers, and children themselves*. Minneapolis: Univer. of Minnesota Press, 1952. Pp. xii+116. \$3.00.
- HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: Univer. of Minnesota Press, 1953. Pp. vi+153. \$3.50.
- HOYLES, J. A. *The treatment of the young delinquent*. New York: Philosophical Library, 1952. Pp. viii+271. \$4.75.
- HUMPHREYS, J. A., & TRAXLER, A. E. *Guidance services*. Chicago: Science Research Associates, 1954. Pp. xvii+438. \$4.75.
- KATZ, B., & LEHNER, G. F. J. *Mental hygiene in modern living*. New York: Ronald Press, 1953. Pp. x+544. \$4.50.
- KNAPP, R. H. *Practical guidance methods*. New York: McGraw-Hill, 1953. Pp. xi+320. \$4.25.
- LINDGREN, H. C. *Psychology of personal and social adjustment*. New York: American Book Co., 1953. Pp. ix+481. \$4.50.
- LITTLE, W., & CHAPMAN, A. L. *Developmental guidance in secondary school*. New York: McGraw-Hill, 1953. Pp. xi+324. \$4.50.
- PEPINSKY, H. B., & PEPINSKY, PAULINE NICHOLS. *Counseling: Theory and practice*. New York: Ronald Press, 1954. Pp. vii+307. \$4.50.
- RECKTENWALD, L. N. *Guidance and counseling*. Washington, D. C.: Catholic Univer. of America Press, 1953. Pp. xiii+192. \$3.25 (\$2.50 paper cover).
- SANDERSON, H. *Basic concepts in vocational guidance*. New York: McGraw-Hill, 1954. Pp. xiii+338. \$4.50.
- SHAW, F. J., & ORT, R. S. *Personal adjustment in the American culture*. New York: Harper, 1953. Pp. ix+388. \$4.00.
- TYLER, LEONA E. *The work of the counselor*. New York: Appleton-Century-Crofts, 1953. Pp. xi+323. \$3.00.
- WARTERS, JANE. *Techniques of counseling*. New York: McGraw-Hill, 1954. Pp. viii+384. \$4.75.

Perhaps the most telling sociological phenomenon in psychology is its rapid professionalization during the decade since the end of World War II. In a scant ten years, an essentially academic discipline has transformed large segments of itself into a form of public service, meeting in the process

the problems with which a service profession must cope: relationships with other professions, the formulation of ethical codes, provisions for internal policing of professional practice, the development of training standards and training opportunities, concerns about legislative recognition, the maintenance of public relations, and many others.

One problem that cuts deeper than others, however, is seldom faced squarely. Psychology's professionalization has occurred largely as a response to a social demand intensified and given justification, although not created, by wartime experience. Such a demand takes primary form as a market for new workers to deal helpfully with the troubles of unhappy people. Since the knowledge necessary to this task has not yet been fully discovered and developed, a need is created for a kind of substitute for knowledge, a literature of practical wisdom and summarized clinical experience. And because the human problems toward which professional psychology is directed are genuinely important and poignant, because something crucial in the lives of his clients is at stake, the professional psychologist himself often is motivated to seize desperately on any technique or idea that has the appearance of usefulness. When the chips are down, as they generally are in professional practice, skepticism about one's own resources is a luxury that few can afford.

One outcome of this state of affairs is the current rift between "pure" and "applied" psychologists. Possibly a bit jealous of the social prominence and income of their colleagues in clinical and counseling positions, research men and serious theorists are likely to judge the literature and oral pronouncements in these fields,

often with some justification, as soft-headed and close to meaningless. On the other hand, made defensive by such charges and harried enough by their applied responsibilities, the professionals tend to reply in hasty irritation that the work of the scientists is unfeeling and irrelevant.

Actually, of course, social need has always outrun available knowledge, generally providing the essential spur to its pursuit. To say that professional practice must rely on hunch and accumulations of uncontrolled experience is to say nothing derogatory *so long as one knows what is happening*. The significance of the present split in psychology is that the element of faction makes it harder for people to acknowledge the difference between knowledge and practical approximations, to work in concert for the replacement of the latter by established fact and soundly based theory, and to formulate from professional practice fruitful hypotheses for the advancement of psychological science.

The professionalization of psychology, then, provides not only opportunities for the discipline's being of service to a needful public but also opportunities for the enrichment of psychological science through the hypothesis-forming potential of applied work. Its danger lies in a splitting of science and profession to the detriment of both. The extent to which this divisive tendency has permeated or been controlled in the field of counseling psychology and adjustment can be estimated from these 18 recent books.

CONCEPTS OF ADJUSTMENT

Willy-nilly, what counselors do in their relationships with clients is determined in part by their implicit or explicit notions of what constitutes

effective personal adjustment and their unformulated or articulate concepts of how it is attained. Most counselors, clinicians, and social psychologists find themselves involved in difficult and ongoing struggles to clarify such questions as the relationship of social conformity to the gratification of individual needs, of articulation with the cultural group to personal spontaneity and novelty, and of the acceptance of social regulation to the promptings of one's individual experience conceptualized as conscience.

Shaw and Ort treat these problems courageously and in a sophisticated fashion in a book that blends effectively clinical experience with research evidence and theories that have had some brush with the laboratory as a proving ground. They argue that the individual's adjustments are best understood as ways of interacting with other people and as products of primarily social experience. This emphasis on one's history and current behavior in interpersonal, social, and cultural contexts is formulated within a conceptual framework of reinforcement learning theory in which personality development and functioning is comprehended as the result of rewards and punishments experienced in the social learning environment.

On the other hand, they are most mindful of the problem of adequacy in personal adjustment and wrestle manfully with the criterion of mere social conformity to which a sociological point of departure often leads. Their resolution of the difficulty is hardly a final one. Shaw and Ort rightly hold that the ability to act in accordance with social standards is an essential part of effective interaction with one's inescapable interpersonal world but that it must be balanced by something they call

"integration." This factor is defined as the maintenance of oneself in one's environment, and behavior disorders are conceived as deviations from adequate integration.

This concept involves two interlocking difficulties. The first is the theoretical embarrassment that derives from the attempt to fit the notion of self-maintenance, deriving from a phenomenological tradition, into the objectivist behavior theory that is the conceptual groundwork for *Personal Adjustment in the American Culture*. This task of unifying different psychological theories is a most worthy one, but the attempt here results in a weakening of the rigor of reinforcement theory and a loss of the intuitive impact that seems to have made phenomenology attractive for many applied workers. Second, the effort to deal with integration apart from conformity lands the authors in the problem of accounting for integrative behavior that runs counter to their social experience, of developing explicit criteria for judging when nonconformity represents adequate adjustment and when it suggests mere rebellion or lack of social awareness, and of providing a *psychological* basis for understanding integrative adjustments that lie outside cultural norms.

Two comments seem in order. The first is that one of the sources of uncertainty and a kind of fuzziness of thought in much of applied psychology may be the necessity of involvement in questions which are essentially ethical or philosophical in character. In a sense, Shaw and Ort accepted the task of defining "the good life" or "the elements of good conduct." A. E. Taylor in his life of Socrates (7) makes much of the Socratic "sense of the importance of implicit obedience to lawful authority" without falling into the "vice of

exalting the mere letter of the law above its spirit." It would seem that Shaw and Ort's courageous book, for all its labors, says little more on this difficult topic of the balance between conformity and integration. There is a temptation to say that psychology should stick to its traditional last, but as May (5) and Mowrer (6) point out, the question of what kind of behavior is associated with human happiness is a legitimate one for science to ask and an inevitable one for the psychological professions.

The second point to be made is that few books illustrate as well as Shaw and Ort's how far a knowledgeable fusion of experimentally based theory and clinically based hunch can achieve two goals. One is the clarification of processes observed in the consulting room and in the field. Defensive operations, the phenomena of age grading and other forms of role taking, masculinity and femininity, and occupational strivings become less complex and communicable in more manageable terms when analyzed according to a theoretical system the strengths and weaknesses of which are assessable in part through laboratory tests. The other is the enrichment of a theoretical structure by seeing how well it can be made to cover practical observations. Even though it was written as a relatively low-level textbook, this volume will repay close reading by those interested in developing research problems in that important area that lies between restricted theory and loose practice.

The volume by Katz and Lehner, on the other hand, is theoretically much less ambitious. Its pretensions lie in the direction of explaining human behavior to unsophisticated readers and of demonstrating to these readers how they may achieve happier adjustments in their own

lives. As such, *Mental Hygiene in Modern Living* is a kind of cross between text and self-help book. This hybridization could conceivably make for certain strengths. For example, the encyclopedic coverage virtually assures most readers of finding something useful in its pages. Like an encyclopedia, however, the book lacks system and thoroughly confuses practical wisdom with verified knowledge and substantiated theory. Nowhere does it come to grips with the problem of what constitutes *positive* adjustment in any cogent way, and the discussion of maladaptive behavior, including the psychoses, leads one to ponder the observation that clinicians and counselors so often seem more at home with psychological ills than with psychological health. Even the technical vocabulary seems wanting in words to describe effective adjustment and the conditions which determine it. It is little wonder, although a trifle alarming, that so little research is done and so little theorizing devoted to the development of personal assets when professional attention is directed so strongly to pathology. One unfortunate result is that discussions of positive personality growth have, as they do in this book, the leaden ring of clichés sounding through them. Saddest of all, this fault cannot be charged directly to the authors. Citing a wide literature, they refer to virtually no studies directly concerned with psychological health; one guesses that they found none. For counseling psychology in particular, with its official emphasis (1) on the positive and the "normal," this absence of careful thought and empirical evidence on the problem of what constitutes normality and its determinants seems to be a deadly realm of ignorance.

Lindgren makes some attempt to deal with this worrisome gap in knowledge and theory in a readable book that pays little attention to the major forms of psychopathology and concentrates on the understanding of "ordinary people." His approach is not discernibly systematic, like Shaw and Ort's, but neither is it a wide-ranging mass of unconnected propositions and unorganized bits of information, like Katz and Lehner's. Refreshingly, Lindgren recognizes human resiliency, the capacity to absorb considerable amounts of frustration without serious damage, as a primary factor in mental health; and he also deserves much credit for bringing *thinking* back to the list of psychological functions of importance to counselors. Dynamic psychology, in its reaction to the rationalist tradition and its proper underscoring of affective and conative determinants of behavior, has forgotten until very recently Freud's (4) remark, "The voice of the intellect is a soft one, but it does not rest until it has gained a hearing." It is to Lindgren's credit that he gives due emphasis to the role of thought in solving common human difficulties without falling back on the notion that the problems of people are always at bottom intellectual affairs.

Nevertheless, the emphasis on pathology that has characterized modern psychology, especially in its professional aspects, seems to provide the undercurrent for this, as for other, volumes. The conception of man implicit here is one that depicts him as being buffeted by his circumstances, protected only by his degree of native toughness and his wisdom in finding "therapeutic" experiences either within or without the consulting room. Everyday social life, work, and religion (conceived naturalistically as a form of ready communication

with a group and its leader) are all discussed as forms of therapy, i.e., as *corrective*. This view may be quite correct, but it is not unchallengeable. One can legitimately raise the question of whether work and personal relationships cannot have positive effects in themselves, not as correctives to unavoidable ills, just as one can ask if the inescapable frustrations and compromises of daily living necessarily require therapeutic relief in mature people. This rather different idea of human functioning might lead to a re-evaluation of experiment and theory and the development of a fruitful new line of psychological research on the nature and antecedents of normality.

One might expect that these concerns for the *nature* of adequate adjustments, involving something very close to ethical and philosophical considerations, might find more explicit attention in two books written from religious points of view. Hoyles' book on delinquency is by a minister of the Church of England, and Recktenwald's discussion of guidance procedures is by a Roman Catholic. The first is a humane and intelligent argument for rehabilitative rather than punitive treatment of juvenile offenders. The second is a workmanlike and comprehensive brief overview of guidance techniques in school settings. Neither contains anything particularly novel, and Hoyles seems more preoccupied with presenting a challenge to the Anglican Church than with analyzing the character of criminal adjustments among young people. The important thing about them which justifies considering them together here is their mutual lack of attention to the problem of what constitutes adequate or integrative behavior. Except for a very few random pages, Recktenwald's book could have been

written by a non-Catholic, and Hoyle uses the language of moral theology to say essentially the same things that anybody would say who conceives of delinquency as a form of behavioral pathology rather than a kind of wilful violation of social tenets. Neither grapples as frankly as do Shaw and Ort, for example, with the problem of adequate adjustment or the behavior patterns associated with "the good life."

RESEARCH STUDIES

If the nature of integrative adjustment is given little attention in these books on mental hygiene, it is not surprising to find a similar lack of concern in the three research monographs by Hathaway and Monachesi, Craig, and Griffiths. The noteworthy thing about these three very different publications is that they represent the vigor with which sound empirical work is being pursued in a professional field. All three, however, have only remote connections with comprehensive theory, which may reflect again the lack of articulation between the scientific and the systematic tradition and the professional one in psychology generally. Their diversity is also worth noting as indicative of the number of problems accepted as the investigative responsibility of counseling psychologists and those concerned with school guidance.

The Hathaway and Monachesi monograph is an exceptional piece of work for two reasons quite apart from the high level of research sophistication that marks it. First, concerned with the prediction of delinquent behavior in children, it avoids the pitfalls of retrospective data. Instead of studying the attributes of those already guilty of delinquent acts as opposed to those of youngsters who are not delinquent, Hathaway and his co-workers began with the

examination of large numbers of subjects below the age at which delinquency tends to occur. Predictions were made at that point, and their accuracy was determined by a close follow-up of the children.

Second, this little book is a splendid exhibition of what can be done with structured personality tests. Using the MMPI, these researchers concerned themselves with profile patterns rather than individual scores, thus introducing a regard for the complexity of test responses similar to that insisted upon by Rorschach workers and the users of other projective techniques. The method of coding profile patterns, however, does not sacrifice the rigor and exactness associated with quantitative scores but usually purchased at the cost of a good deal of information about the respondents. In short, the methodological implications of this study are that objective personality scales can be used in such a way as to retain precision without exorbitant cost in what Cronbach (2) has called "bandwidth." In a field where assessment procedures and predictive devices often yield blurred and fuzzy results, the model provided here assumes particular importance.

As for findings, Hathaway and Monachesi present convincing data to indicate that adult test patterns of the amoral psychopath and the hypomanic individual tend to occur with predictively significant frequency among predelinquents, whereas the occurrence of test patterns characteristic of adult neurosis seem to exercise an inhibiting influence on the development of delinquent behavior. Similarly, those test patterns which show no high deviations and are generally considered as "normal" are remarkably predictive of nondelinquent adjustments. This study

strongly suggests that the interpretation of delinquency as a form of neurotic behavior overlooks fundamental differences in neurotic and criminal adjustments and implies that the dynamics of delinquency are not to be found in repressions and self-derogating tendencies.

On the other hand, there is little in this monograph to shed much light on the antecedents of delinquency or its modification. Meaningful investigations of the experiential determinants of delinquent behavior, of the relevant learning environments in which delinquent patterns are acquired, are yet to come. Until such inquiries are undertaken with the same vigor, sophistication, and insistence on exactness that one finds in this prognostic study, it is unlikely that a useful theory of delinquent adjustment will be developed to replace current clichés, hunches, or scattered bits of knowledge.

Griffiths' study of children's behavior difficulties is not concerned with delinquency. Rather, he was concerned with the kinds of behavior problems that children themselves think they have. Studying youngsters from six to fourteen, he finds that awareness and acceptance of social rules and regulations increase with age, that younger children tend to be aware of overt aggression as a source of trouble but only later to develop a sense of uneasiness about submissive and withdrawing tendencies in themselves, that children generally are surprisingly aware of how both parents and teachers would like them to change their behavior, and that children from the middle socioeconomic group are more conforming to adult norms than those from other classes.

The most outstanding thing about this study is its documentation of the effects of socialization. As age—

which means social experience—increases, children become increasingly sensitive to adult demands and norms of conduct and even at relatively early levels show surprising insight into adult desires to have them change their behavior in certain identifiable ways. Moreover, children seem to discriminate fairly early the different demands made of them by different adults. Fathers, for example, are perceived as persons who primarily ask that they not be disturbed. Teachers are people predominantly interested in an orderly and efficient classroom, although youngsters seem a bit more aware of the interest of teachers (in contrast to parents) in the modification of withdrawing and submissive traits. Mothers, like teachers, want smooth-running homes and evaluate children's behavior accordingly.

Methodologically simple, this monograph provides a wealth of information about children's evaluations of their own behavior and relevant to an increased understanding of the socialization process. Again, however, the findings are not related to any generative kind of theory that would lead readily to more systematic research on the socialization process or the antecedents of children's attitudes. The importance of this problem also requires comprehensive study to deal with such questions as the influence of perceived adult norms on peer relationships, the development of conscience, and the development of personality. For example, does the child who becomes aware of adult conduct norms at an early age tend to develop a greater degree of responsiveness to authority than the youngster who acquires this kind of awareness later? To what extent is this kind of awareness related to the type of rewards and punishments typical of the home

and to the pattern of child-parent relationships generally? Griffiths' study provides an excellent point of departure for theoretically important investigations of the antecedents of all kinds of adjustment problems, including that of "adequate" or integrative adjustment patterns.

Craig's monograph on guided learning is in the tradition of transfer-of-training experiments of the Thorndikian type, but its implications are somewhat broader than most such studies. Working with recent college graduates as subjects and using a task consisting of multiple-choice verbal test items, Craig found that giving learners a short statement of principles common to a group of items reduced errors, increased the efficiency of solutions, and aided the process of discovering the basis for correct response to other items. This effect became more pronounced as a function of the difficulty of the learning situations. While conceived within narrow and well-controlled laboratory conditions, the experiment suggests that interpersonal as well as intellectual problem solution might well be facilitated by the development of principles common to many situations through participation in discussion groups, orientation classes, and individual guidance conferences. Growing out of the tradition of law-of-effect learning studies, Craig's inquiry has implications that amount to ready-made hypotheses for counseling psychologists to put to test. This kind of research would have the considerable advantage of being tied to a comprehensive theory of behavior that could be expanded in two directions: application in a knowledgeable kind of way to significant social problems and theoretical extension to affective and conative spheres.

COUNSELING TECHNIQUE

The great bulk of these books is devoted to technique, the methods to be employed by the practical counselor. This state of affairs is perhaps a function of many factors. One is the historical emphasis on practicality and tangible action in American culture, one current manifestation of which is the spate of how-to-do-it books in many areas. Another is the intensity of the demand out of which the professionalization of psychology has grown. With a greater awareness and acknowledgment of human problems, there is a greater insistence that preventive and remedial treatment be supplied by *somebody*. One outcome seems to be a crop of manuals and books of instructions for those who are pulled into this vacuum of need, often without the fundamental training upon which professional practice must rest.

The striking thing about the present collection of nine volumes is their similarity and degree of overlap in content. The books by Arbuckle, Humphreys and Traxler, Knapp, Little and Chapman, Recktenwald, and Warters all say essentially the same things in essentially the same ways. One marvels at a market that can absorb such comparable publications. It is not that these books are lacking in workmanship or competence; they have both qualities. But they are all very much in the same vein of describing activities by which a counselor can keep himself busy doing technical things.

Of these technical things, two are given major predominance. One is testing, the other is record keeping. Both, of course, are important, but one wonders if lists of psychometric instruments and lengthy discussions of cumulative records are the core of the counseling process. Counselors

certainly need methods of sharpening the precision of their observations and ways of estimating how a given client or student compares with his fellows. Tests are very relevant. Likewise, counselors must know how to keep account of the development and progress of those to whom they are responsible. Record-keeping devices are very much in point. The present manuals, however, devote very little space to the ways in which these technical aids can be used to advance the aim of the counseling process, the *helpful modification of client behavior*. Neither is there much space afforded to the use of the data collected by such instruments in the advancement of knowledge. If counseling psychology is to remain a part of *psychology* and not to degenerate into mere technician status, then surely research and careful thought about human problems as they are encountered on campuses, in industry, and in the consulting room generally must be willingly assumed functions of practitioners of the counselor's craft.

If this point can be regarded as debatable, however, another can hardly be. The essence of the counselor's job is the modification of client behavior through face-to-face contacts. It is *counseling*. The proportion of pages in six fat volumes concerned directly with this face-to-face process of behavior modification is of the order of eight per cent! One suspects that under professional pressure it is somehow easier to be occupied with testing programs and the maintenance of records than with direct personal service. One also suspects that the lack of real knowledge about the counseling process is an overwhelming if inarticulate determinant of this strange fact. Is it possible that the how-to-do-it man-

uals are premature in many ways and that more research and more thought on the counseling process and its outcomes are prerequisites yet to be achieved for preparing truly meaningful books on technique? The profession is young, and its undergirding of basic knowledge may be too weak still to support such a structure of textbooks. The same comments apply to Arsenian's little book, a set of instructions for YMCA secretaries who find themselves involved in counseling relationships.

Two books are exceptions. One is Driver's discussion of the technique of multiple counseling, the other Tyler's sophisticated and informed *The Work of the Counselor*. The first is an extensive and interesting presentation of small-group discussion methods of modifying behavior drawn primarily from a fusion of client-centered psychotherapy and group dynamics. The book is noteworthy because of its extensive case excerpts, its serious if rather loose attempts to evaluate procedures and outcomes, and its step-by-step descriptions of multiple counseling groups and their development. Anyone interested in group counseling methods will find rewarding reading here, and the challenge to rationalize the method and to subject it to rigorous investigation is considerable. Research hypotheses abound in its pages, and the book's freedom from a doctrinaire tone invites the test of empirical inquiry.

Tyler's volume, by far the best of the "practical" books on the present list, frankly acknowledges the status of counseling as an "art," dependent more on the accumulation of shared clinical experience than on research evidence and tightly constructed theory. Nevertheless, she argues that

the rapidity of improvement in counseling procedures must rely on the soundness of relevant investigation and that intuition can be made articulate through research. Her chapters, therefore, are organized in terms of what counselors appear to do and think in their practice but are supplemented with research summaries reviewing the relevant research literature and clarifying issues on which research may throw some light. Unlike other authors, Tyler's greatest emphasis is on the interview, and her understanding of the relationship between observation on the one hand and behavioral modification on the other is insightful and seminal while appropriately humble.

Oddly, this fine book pays virtually no attention to personality dynamics or the problems of clarifying the concept of adjustment. This is the more remarkable since Tyler's is the only one of these volumes that makes it explicitly clear that the profession of counseling cannot be successfully followed by one whose only equipment is a bag of tricks, a knowledge of available tests, and ways of keeping records. Much of what she says implies that counselors require more than anything else an ever-developing knowledge of learning, perception, motivation, and human development, and the failure of her book to concern itself with these substantive issues in some degree is surprising and disappointing. There is a limit, however, to which an author can be taken to task for not writing a different book, and Tyler's contribution is too informed and too well based on general psychological knowledge and in the traditions of psychological science to permit much legitimate fault finding.

THEORETICAL CONTRIBUTIONS

As a corrective to the how-to-do-it manual, professional psychology has two major theoretical alternatives. It can attempt to systematize its observations and to develop a formal body of propositions independently of the rest of psychological science, risking a fragmentation and divisive force that might well deprive the profession of the intellectual and scientific bases on which most mature professions rest. Or, it can formalize its observations in terms of general psychological theory, a process which demands the assumption that such processes as learning, motivation, and perception are essentially similar whether studied in the laboratory or in the counselor's office.

The little book by Pepinsky and Pepinsky is an outstanding example of the latter possibility. With even more cogency and creativeness than Dollard and Miller's *Personality and Psychotherapy* (3), this work illustrates the advantages of relating clinical experience to a formal and elegant theory derived from experiment and the theorist's study. The Pepinskys both apply and enlarge a relatively rigorous Hullian model by using it to account for anxiety and unproductive behavior as these concepts are employed descriptively by counseling practitioners. With a minimum of concern for specific techniques, *Counseling: Theory and Practice* demonstrates how a counselor can better understand and communicate with his clients if he is thoroughly in possession of a comprehensive theory that is well fortified by careful research than if he attempts to rely on a bag of technical tricks or the unrelated and some-

times inconsistent assertions that occasionally pass for theory among those whose training is deficient in general psychology.

Indeed, a major thesis of the book is that professional practitioners must work *both* as applied clinicians and as scientists, each role facilitating the other. Experience as a counselor provides observations that allow the inductive formulation of general laws. Experience as a scientist permits the testing of these laws and the determination of which are useful, which require modification, and which must be scrapped as unproductive of either new insights into the counseling process or new hypotheses for research. Practice unleavened by theoretically oriented research responsibility leads to a softness of principle and a blurring of theoretical precision that is likely to reduce counseling to a matter of *mystique* and uncommunicable intuitions. Theory-oriented research without familiarity with practical problems and clinical events risks a narrowing and a rarefying that make theory lose significance and, worse, lose comprehensiveness.

Sanderson's *Basic Concepts in Vocational Guidance* is a very different matter, more illustrative of the first alternative, the construction of theory from professional observations without regard to more general psychological propositions. It is possible to read this volume without being reminded that human beings learn and perceive and manipulate symbols in thought or social communication! And yet the book is a good one, infused with the attitudes of science if not its general psychological content and reflecting a careful and thoughtful evaluation of a wealth of litera-

ture and counseling experience. Contrary to the how-to-do-it books, this one rests on the proposition that counselors rely too heavily on "diagnosis" (testing and record keeping), and too little on "the helping process itself" (behavioral change through face-to-face contact). Sanderson shares with the Pepinskys, then, the belief that counseling practice is facilitated more by a coherent body of ideas by which the clinician can understand his clients and his procedures than by a clutch of mechanical methods.

Further, his conception of the counseling role is a broad one, taking fully into account the motives and affects that revolve around the status aspects of occupational adjustment and the variety of noneconomic rewards bound up with work. But this very breadth of conception leads to disappointment when the reader finds nothing about the general psychology of personality or the research evidence and tentative principles formulated from it that bear on personal development, the making of decisions, or the socially acquired drives that determine so much of human activity. The alternative of reasonably precise but discreet and unsystematized concepts drawn from counseling experience itself may help to clarify the counselor's present work, just as Sanderson's stress on understanding the helping process as contrasted with reliance on technique may help to increase the actual amount of service given to clients. It is most unlikely, however, to advance the field, to stimulate a search for new knowledge of relevance to the process or its outcomes, or to increase the understanding of those general human functions which are essentially the

subject of the counselor's concern, just as they are the subject of interest of friends, experimenters, and virtually all others who find the study of man inevitable as well as fascinating.

Neither of these two competent books deals with the problem that seems so central even while it is so thoroughly ignored. What constitutes effective adjustment or integrative behavior? Both Sanderson and the Pepinskys are essentially silent on the question, although the latter skirmish with it briefly in a provocative chapter on the assessment of counseling processes and outcomes. It seems somehow unlikely that fruitful outcome-research can be accomplished until this basic element in establishing a criterion is thought through.

SOME GENERAL COMMENTS

If the present list of books is a fair sample of the present state of counseling psychology, there is room for worry over the effects of rapid professionalization on psychology as science. With the public demand for service high, there is apparently a considerable interest in publications dealing with *techniques*, less in books seriously occupied with *knowledge and understanding*. There seems to be a tendency for practice to divorce itself from science that could be detrimental to both. Should counselors become only technicians and general

psychologists only runners of rats and manipulators of brass instruments, both psychology and a needful public will suffer. The professionalization of large elements in psychology is welcome not only because it provides an opportunity to serve troubled people but because it opens new vistas for psychological research and the understanding of behavior. The problems are large enough so that their solutions probably lie in the collaboration of practitioner and rat man or tachistoscope operator. Without such collaboration, productive solutions seem far distant.

But there is also evidence to justify ending this review on other than a Cassandra note. The books by the Pepinskys and Shaw and Ort and the three research monographs all indicate an investigative vigor and a retained relationship with general psychology that holds much in the way of theoretical promise, just as Tyler's volume and, to a lesser degree, Sanderson's suggest that careful thought, a proper appreciation of research, and a knowledge of available evidence from the laboratory and the field still have a stout influence on practice. Division is by no means inevitable, and perhaps only a little awareness of the issues is necessary to insure the development of a useful profession firmly rooted in psychological knowledge and rigorously developed theory.

REFERENCES

1. AMERICAN PSYCHOLOGICAL ASSOCIATION, DIVISION OF COUNSELING AND GUIDANCE, COMMITTEE ON COUNSELOR TRAINING. Recommended standards for training counseling psychologists at the doctorate level. *Amer. Psychologist*, 1952, 7, 175-181.
2. CRONBACH, L. J. Report on a psychometric mission to Clinica. *Psychometrika*, 1954, 19, 263-270.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. FREUD, S. *The future of an illusion*. New York: Liveright, 1928.
5. MAY, R. *Man's search for himself*. New York: Norton, 1953.
6. MOWRER, O. H. Neurosis and psychotherapy as interpersonal processes: a synopsis. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald, 1953. Pp. 69-94.
7. TAYLOR, A. E. *Socrates*. Garden City, N. Y.: Doubleday, 1953.

BOOK REVIEWS

CARMICHAEL, LEONARD. (Ed.) *Manual of child psychology*. (2nd Ed.) New York: Wiley, 1954. Pp. ix+1295. \$12.00

The second edition of the *Manual* presents an excellent picture of child psychology. In this respect it is a better book than the first edition, which in the reviewer's opinion was out of date when it was issued.¹ This does not mean that this is a definitive work. It is, rather, an expansive, factious book. But it would be surprising if it were otherwise, considering the unfinished state of the frontier area with which it deals. At the same time it is an ambitious book that faithfully reflects the vigor and aspirations of this youthful science. The title of the volume, itself, expresses the ambitious tone of much of the book. A manual, by dictionary definition and by common usage, is a book of reference which can be conveniently carried in the hand. But this book's four and one-half pounds make it too big to be easily handled, and it is a book to be studied and argued with, not one to be referred to on the run for authoritative answers to standard problems.

The book is similar in organization and authorship to the first edition, seventeen of its nineteen chapters having the same titles and sixteen the same authors as the first edition (two chapters have different junior authors). Chapter changes from the first to the second editions are: "The Feeble-minded Child" (Doll) is incorporated into a more inclusive chapter called "Psychopathology of Childhood" (Benda); "Maturation of Be-

havior" (McGraw) is omitted; "Social Development" (H. H. and G. L. Anderson) is added; "Adolescence" is written by Horrocks in the place of Dennis. A number of chapters in the new edition reprint large sections from the previous one in unchanged form. To be exact, 70 per cent or more of the following chapters are re-printings of the first edition: "The Onset and Early Development of Behavior" (Carmichael), "Animal Infancy" (Cruikshank), "The Neonate" (Pratt), "Physical Growth" (Thompson), "The Ontogenesis of Infant Behavior" (Gesell), "Learning in Children" (Munn), "Mental Growth" (Goodenough), "Research on Primitive Children" (Mead), "Behavior and Development as a Function of the Total Situation" (Lewin, with a supplement by Escalona), and "Gifted Children" (Cox). About half of this edition of the *Manual* is a reprinting of the first edition, and about half involves major revisions or new topics.

The ten chapters listed above will not be considered in detail here. The reader is referred to the earlier review for comments on them, as their analyses of the topics under consideration are essentially unchanged. These are research areas which the authors see as having been on a plateau in recent years so far as new developments are concerned. This does not mean, of course, that important additions, replications, and verifications have not occurred. For example, a keystone in the study of gifted children was put in place during this period by Terman and Oden's twenty-five year report.

The reviewer must repeat his objection to the chapter on the physical growth of children. He agrees

¹ Roger G. Barker. Manual of child psychology, a special review. *Psychol. Bull.*, 1947, 44, 162-170.

completely with the author that "child behavior cannot be understood apart from . . . the physical body through which it perceives, reacts, and functions, and to which others react," (p. 292); but unfortunately he finds here no consideration of this problem. Can't we have a chapter on the *psychology* of physical differences in children in the next edition such as we have on the psychology of sex differences in this one?

The chapters of the *Manual* where major changes appear, and the new chapters, may be individually mentioned as follows:

"Methods of Child Psychology" (John Anderson). Anderson wrote chapters with this title in the 1933 edition of Murchison's *Handbook* and the 1946 edition of the *Manual*. These, with the present essay, constitute a revealing exhibit of the rewards of two decades of preoccupation by psychologists with the problems of methodology. Anderson wrote approximately 11,000 words in 1933, 22,000 in 1946, and 33,000 in 1954 and while the increase in level may not be quite so marked as the increase in quantity, the reviewer perceives it as being nearly so. However, the scope of the chapter is too wide; it is, in fact, a survey of general psychological methodology. With a few changes in nomenclature and examples this essay would be as appropriate in a social or an experimental psychology as it is here. One hopes that the time will come when the general field will have been so well covered by others that Anderson can devote his attention to *special* problems of methodology in psychological research with children.

"The Environment and Mental Development" (H. E. Jones). Jones' review shows that there has been an active and profitable period of work on this area. He reports a welcome

movement away from global nature-nurture studies to increased specification of the effects of particular environmental influences. Jones appears to take the position, without elaborating and emphasizing it sufficiently, however, that this movement must go still further. He says in connection with the Iowa studies "our chief need in this field is not for statistical methods of greater power and subtlety but for more vigorous experimental procedures . . ." (681). If he means here, as the reviewer thinks he does, the need for better definition and control of environmental variables, he has raised a crucial point. In its context, Jones' statement seems to imply that, for example, it is not enough to study the effects of nursery school attendance upon intellectual development; it is necessary to define the significance of nursery school for children in psychologically meaningful concepts: perhaps in such variables as the stimulation, the attractiveness, the freedom, the rewards, and the learning opportunities of the nursery school situation. This points to an important requirement which Jones' review does little to meet. Although his critical evaluations of the technical adequacy of individual studies are excellent, he does little to organize and conceptualize the kinds of problems in this field. All "environmental" influences, ranging from culture, economics, social class, foster homes, hospitalization, and schooling to birth order, season of the year, nutrition, health, physical size, physiological maturity, and race are treated on the same level. Surely the time will soon come when some discrimination and sorting of problems within this field will be possible in terms of primitive theories of environmental variables.

"Character Development in Children—An Objective Approach (Ver-

non Jones). The author reports many new investigations of relevance to this topic, but the area of understanding appears to have been increased very little in the time since the first edition. It is clear that this is an area where the multiplying of facts is proving of little benefit. New directions are needed.

"Language Development in Children" (McCarthy). This is the longest chapter in the *Manual* (120 pages) with the largest number of references (776). It has been reorganized and greatly expanded for this edition. Extensive new sections cover the vocalization of infants and language disorders in relation to personality development. The problems considered cover a tremendous range including, for example, fetal sounds and the birth cry; vowel and consonant sounds in infant vocalization; vocabulary tests; amount and rate of talking; the function of language in the child's life; the effects of institutionalization and multiple birth upon language development; the correlation between language and motor, intellectual, and social development; language disorder syndromes and personality development; delayed speech; articulatory defects, stuttering. Despite McCarthy's very effective organization and condensation (including two four-page summarizing tables), the amount of material is so overwhelming that one leaves this chapter with a great weariness. Perhaps the basic difficulty is that "spoken language" is in itself not a useful category of behavior; perhaps it is only phenotypic, symptomatic behavior like hand behavior and foot behavior. In any case, to make future reviews more effective, some means of focusing upon the psychologically most meaningful and timely issues is essential. As it stands, this chapter provides a useful guide to a

vast literature.

"Psychological Sex Differences" (Terman and Tyler). This chapter has been largely rewritten for this edition of the *Manual*, and it might well serve as a model for the kind of exposition suited to this book: the problem of the chapter is clear, the evidence is not beyond the limits of what can be handled in the space available, the authors make explicit the limits of their review and of the bibliographical sources covered, they include only those publications most relevant to the trend of the evidence as they see it. This method reveals clearly the status of the problems as viewed by the authors. Developments in the interval since the first edition of the *Manual* are more adequate data which confirm earlier findings and which reinforce the authors' earlier caution regarding possible biological and cultural roots of psychological sex differences. Theories and speculations about the possible lines of influence of these factors are not covered. This is a feature of the problem which some might consider of value.

"Emotional Development" (Jersild). This is a completely reorganized and greatly expanded survey of emotionality in children. Jersild places his discussion within an explicitly stated conception of emotion. He considers not only research findings, but various theories and speculations as well. He covers not only the well tilled fields, but also such newly cleared research areas as joy and compassion. It is beyond the scope of the present review to evaluate this chapter in detail. While it can doubtless be criticized from many viewpoints, it is effective as a structure for presenting much of the current state of knowledge and thinking in this area, much more so than the chapter in the previous edition.

"The Adolescent" (Horrocks). It is difficult to conceive of two chapters with this title more different than those of Dennis in 1946 and Horrocks in 1954. Some measure of this difference is found in the fact that while between them they refer to 337 citations of relevant literature dated 1941 or earlier, only 12 per cent of these citations are common to both bibliographies. This must carry an important lesson regarding the preciseness with which *adolescent* is defined and the "reliability" of these two "observers" of the relevant literature. While Dennis defined his problem as that of the behavioral correlates of physiological sexual maturation in humans, Horrocks ranges much more widely. He devotes six pages to the history of studies of adolescence, thirteen pages to the physical and physiological aspects of adolescence, and nine pages to behavior in adolescence. This is a rather elementary, didactic essay on many aspects of adolescent problems.

"Psychopathology of Childhood" (Benda). This is a new chapter in the *Manual*; it covers mental deficiency (30 pages), personality disorders on a higher integral level (eight pages), and childhood schizophrenia (five pages); it has ninety-four citations in the bibliography. It is beyond the reviewer's competence to make evaluative comments on this chapter.

"Social Development" (H. H. and G. L. Anderson). Harold Anderson has, over a number of years, produced an important series of investigations of the social behavior of children. Here for the first time is a vigorous and invigorating statement of the theoretical framework of this research. It is a theory of almost cosmic scope. A sample of the headings used by the Andersons includes: The Characteristics of Biological Growth; Growth of Civilizations; The Place of

Value; Social Development and the Second Law of Thermodynamics; Psychological Entropy and Culture; Theory of Probability; "Scientific" Prediction and Organization; Disintegration, Deterioration, and Psychosis; The Growth Circle; The Vicious Circle; Friendliness, Cooperation, and Integrative Behavior; Leadership; Social Learning. The bibliography refers to such diverse scholars as Toynbee, Wiener, Sinnott, Conant, Freud, Spranger, Krech, Dennis, Sears, and Schneirla. The chapter begins with an outline of trends in psychological thinking rooted in Darwin, Preyer, and William James and ends with the recent investigations of Lippitt, Bowlby, and Bronfenbrenner. A theory of such scope is beyond the capacity of this reviewer to evaluate, and the task is not made easier by the fact that forty-five pages are too few to cover all the details and make all the transitions. However, this reader found intriguing the world view so badly sketched. Here is the kind of broad perspective child psychology hopes some day to achieve, and this may be the direction the quest will take. It is to be hoped that this essay can be elaborated elsewhere as it deserves.

The present volume exhibits the same ambiguities as the earlier one regarding its functions. Some chapters are written as for a text; others are scholarly contributions with no concessions to immature students; still others are literature reviews similar to those of the *Psychological Bulletin*. Partly as a consequence of these different conceptions the chapters are uneven in style and level of writing. But this variety is also, without doubt, a product of the uneven development of the different areas. Child psychology is, indeed, in a ragged, unfinished state and the *Manual* reflects this in all its aspects.

One instance of this is found in the way the same material is handled in chapters with overlapping subject matter. When the same issue is considered by more than one author, it is not uncommon to find different interpretations of the data. Another instance is found in the treatment accorded two modern giants of child psychology: Freud and Piaget. Several authors make a valiant effort to include psychoanalytic findings. It is obvious, however, that all are uncomfortable in their efforts to be simultaneously tolerant and scientific. Psychoanalysis in this book is like a foreign body. (One cannot but remark that in this respect the second edition is far ahead of the first, where Freud received little mention, and that the *Manual* is far ahead of most books under a psychoanalytical aegis which give academic psychology even less attention.) Piaget, too, is treated as an alien, for reasons less obvious to the reviewer. There is no reference to Piaget's important recent publications.

This is a good picture of child psychology in 1954. The weaknesses of the *Manual* are largely the weaknesses of the science it surveys. A good manual of child psychology awaits a more mature science of child behavior. In the meantime this book and, it is to be hoped, its future editions provide an important aid in achieving this maturity.

ROGER G. BARKER

University of Kansas

THRALL, R. M., COOMBS, C. H., & Davis, R. L. (Eds.) *Decision processes*. New York: John Wiley, 1954. Pp. viii+332. \$5.00.

A summer seminar in 1952 brought together economists, mathematicians, psychologists, and a few representatives of other fields, all concerned with some aspect of choice

behavior. Each participant communicated the ideas in his field which he regarded as likely to benefit the others, and some of them were stimulated by the interchange to start new research. The 19 heterogeneous papers comprising *Decision Processes* were prepared during the subsequent year. The book makes evident the difficulties in this area and discourages any expectation of early pay-off for the psychologist.

We may contrast two broad lines of interest in decision analysis, the normative and the descriptive. The normative, found in mathematics, statistics, and economics, tries to state how a rational being ought to act. The descriptive is found in psychology, and to an increasing extent in economics. Descriptive studies examine actual choices, and seek a law to predict the choices or a rationalization to "explain" them. Models from the normative studies aid in rationalization, and the descriptions demonstrate, among other things, how far normal behavior is from that of the postulated rational being.

Among the normative papers here, one series compares the various decision criteria (e.g., minimax) which are currently under study. A second series deals with the development, from axioms, of utility functions to describe individual preferences. These papers contain some new mathematical developments. Of most general interest is Marschak's "Towards an Economic Theory of Organization and Information" which deals comprehensively with the value of information in making decisions. This paper is not as useful an introduction, however, as Marschak's chapter in *Mathematical Thinking in the Behavioral Sciences* (Glencoe: Free Press, 1954, P. F. Lazarsfeld, ed.).

On the descriptive side, we have a

variety of experiments on preferences among wagers, on probability learning, and on coalition-forming behavior. The prize paper in this group is a delightfully designed experiment by Hoffman, Lawrence, and Fester, demonstrating that coalition forming depends not only on the ostensible pay-offs to be obtained, but also on subtle social attitudes among the participants. Bush, Mosteller, and Thompson present a complicated stochastic model for learning in choice situations. Coombs, Raiffa, and Thrall write helpfully on the place of mathematical models in scientific reasoning, and Coombs employs his "ordered metric" in papers on social choice and (with Beardslee) on wagering.

This volume seems more a memorandum among the participants than a presentation for other readers. These reports do not represent substantial, consolidated advances, as all concerned recognize. There is much use of the single-case experiment, or of the tentative mathematical formulation involving assumptions unsatisfactory to the author. The papers are difficult to read, as mathematics and as English. Even Estes, one of the more lucid contributors, writes this appalling sentence: "When the model is interpreted in terms of the present experiment, it turns out that the rate of learning (systematic change in probability of making a given prediction) depends upon the characteristics of the momentary environmental situation but that over a considerable series of trials the probability of making a given prediction tends to a stable asymptotic distribution with the asymptotic mean, for a group of similar individuals run under like conditions, being independent of the momentary environmental situation, in the present experiment, the nature of

the signal, S."

Many psychologists are likely to learn something from a few chapters which are closely related to their interests, and to which they can bring considerable background. For the reader wishing a general knowledge of the frontier areas represented here, one would recommend parts of the Lazarsfeld symposium, the recent *Bulletin* paper by Ward Edwards, and Irwin Bross' *Design for Decision* (Macmillan, 1954).

The eventual impact of utility theory on psychology may well be considerable, even though investigators are not yet successful in integrating the field. The concept of preference for various outcome distributions intrudes into the interpretation of learning experiments, where psychologists have generally regarded each reinforcement as having an objective value. Utility theory shows that concepts like *level of aspiration* are oversimplified, and the coalition experiments introduce new and provocative problems for social psychologists. The contrast between actual choice and rational choice has many implications for the psychology of thinking, especially as regards its social determinants. Talented investigators have started to mine in what is evidently a good spot, even if this first load contains little pay dirt.

LEE J. CRONBACH

University of Illinois

MOSES, PAUL J. *The voice of neurosis*. New York: Grune & Stratton, 1954. Pp. v+131. \$4.00.

In this volume, the author focuses on attributes of voice, not on verbal communication or speech in the usual sense. Resonance, melody, register illustrate the elements emphasized. Not sharply organized or sufficiently substantiated to serve as a firm introduction to the method or as a con-

vincing summary of research, the book nevertheless opens up intriguing potentialities. Voice patterns reflect peculiarly significant aspects of personality; the neurotic becomes a different voice instrument from the schizophrenic; and therapy combining voice and dynamic elements is indeed a fascinating approach. These are the three themes that run through the book.

The major thesis presents voice as an expressive technique. Just as a graphologist uses handwriting, so Moses uses voice in all its varied aspects to understand personality. He compares a blind analysis with a Rorschach, and integrates other analyses with case material. His ideas about sound approaches to interpretation hold interest for those who use projective or expressive methods. The aspects of voice he emphasizes and his guide lines in analysis suggest that he may have hit on one of the potentially most fruitful areas of expressive behavior—with elements subtly modified throughout important stages of life history and remaining identifiably fixed in adult vocal behavior. With ever increasing fidelity in recording, voice may become a major focus in the understanding of personality.

ROY M. HAMLIN

*Western Psychiatric Institute
University of Pittsburgh*

NUTTIN, JOSEPH. *Tâche réussite et échec: théorie de la conduite humaine*. Amsterdam: Publications Universitaires de Louvain, 1953. Pp. x+530. 330 fr.

Over a period of about fifteen years, the author of this study has been conducting experiments at the University of Louvain that are intended to show the impact of success and failure upon human personality.

They comprise rather simple experiments in perception, learning, and memory. In some experiments, subjects are asked to judge the area of geometrical surfaces, or the number of people shown in pictures. In others, they are asked to recall whether shown two places numbers were added to or subtracted from another such number in a previous showing, or whether words were learned as associated with another word or a number. Arranged in series, the subject is required to render his judgment as each member in the series is presented, and then is told whether his response is right or wrong. In some series the number of successful and unsuccessful judgments are equal, in others one or the other variety predominates. Subsequently, the subject is questioned regarding matters for which the instructions provided no mental set. For the most part, subjects are young people of school and college age. The results of normals are compared with those of manic and melancholic subjects. The experiments are so designed as to be safeguarded against the usual types of psychological error.

Experiments on learning in which some terms are remembered and others forgotten inevitably involve the question as to how successful responses tend to be reinforced. In this context, Nuttin undertakes a critical examination of the meaning of the law of effect. He concludes that Thorndike's view, at least as set forth in his later writings, is too mechanistic. Hull's position is objected to on similar grounds. With respect to the views of Tolman, he is more tolerant. Nuttin expounds a dynamic conception which admits of a cognitive factor in the stimulus-response sequence, and stresses perfection of response rather than repetition of behavior in

the learning situation. Reduction of need is designated the essential condition for reinforcement. He finds support for this conclusion in the recent experiments concerning unfinished tasks.

To this reviewer it seems that, having performed numerous experiments on the effect of success and failure upon incidental memory in simple laboratory tasks, Nuttin has undertaken to widen the scope of the bearing of the experiments in such a way as to provide a general discussion of personality dynamics. In an introductory section under the title "Conduct and Result," he gives a critical survey of the chief contributors to experimental and clinical theory in psychology. Here he develops his theory of personality dynamics in which the cumulative effects of success and failure are often so devastating as to leave traumatic effects.

While those who subscribe to strict objectivity in psychology will be apt to deem Nuttin's appeal to a cognitive factor in behavior dynamics as a regression, those who are disposed to stress the bearing of psychology upon everyday life will find much to acclaim. Left as an open question, however, is the degree to which simple laboratory tasks, such as those in which a subject estimates the number of square centimeters in a triangle or the number of people in a picture and is told he is wrong, approximate such real-life failures as loss of academic standing, rejection by an attractive person of the opposite sex, or elimination from the team in sports.

Using a simple style that is free of the more involved forms of idiomatic expression, Nuttin easily conveys not only his ideas but also his enthusiasm for experiment. By integrating both experimental and clinical sources, the work is given a broad practical bearing which will be of interest primarily

to specialists in learning and in personality development

MICHAEL J. ZIGLER

Wellesley College

KORNHAUSER, ARTHUR, DUBIN, ROBERT, & ROSS, ARTHUR M. (Eds.) *Industrial conflict*. New York: McGraw-Hill, 1954. Pp. vii + 551. \$6.00.

The subject of industrial conflict is so broad that one would hope for its treatment to be lengthy, interdisciplinary, and to encompass many specific topics and viewpoints. As such, this book, written under the sponsorship of SPSSI, meets all one's expectations. It sets out with the ambitious purposes of analyzing the determining factors and conditions which give rise to industrial conflict and of assessing various efforts of solution.

The book is divided into five main parts: (a) Basic issues concerning industrial conflict. (b) Roots of industrial conflict (motivational analysis, organization and leadership of groups in conflict, social and economic influences). (c) Dealing with industrial conflict (accommodating to conflict, efforts to remove sources of conflict, social control of industrial conflict). (d) Industrial conflict in other societies. (e) Industrial conflict; present and future.

Thirty-nine different authors, including academicians, writers, labor leaders, and industrial representatives, have contributed to the book's forty chapters. With such a heterogeneous group, it is inevitable that varying opinions will be found. However, this adds to the book's value by providing a variety of viewpoints and by highlighting some of the unsolved issues. It is exciting to see this interdisciplinary approach being applied increasingly to industrial problems. Certainly such an approach must be satisfying to all "problem-oriented"

industrial psychologists who have long recognized that their medicine bag does not always contain the optimum solution, that many of their tasks are not purely psychological in nature but transverse numerous disciplines.

That the book covers a timely topic is obvious—the tables showing man-days idle as a result of strikes are in themselves sufficient proof. That the book asks more questions than it gives answers, that it presents more suggestions for further research than conclusions from past research, is also true. But it also offers a comprehensive picture for the "serious-minded public" who want a general overview of the relations between labor and management groups and among individuals in the industrial setting.

A final word about the authors. In these days when an increasing number of "edited" books turn out to be reprints of journal articles plus a few introductory remarks, it is refreshing to find a book where each author has not only contributed something new to the field but has done so with no monetary remuneration—all proceeds from the book go to the SPSSI.

JEROME H. ELY

Dunlap and Associates, Inc.

SCHAFFER, ROY. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954. Pp. xiv + 446. \$8.75.

Skinner teaches an alert pigeon to peck a bulls-eye in five minutes, by first reinforcing any approximate success. Struggling with a much more challenging puzzle, the book reviewed here is no bulls-eye, but may be hailed with cautious enthusiasm as the most encouraging near miss of its kind yet published.

The author presents a detailed attempt to establish feedback between

painstaking observation of complex individual behavior and general "laws." The fact that some of the laws are tentative or dubious need not be emphasized with undue distress. The attempt itself hits at the core of the projective problem. The Rorschach was never a test in the Binet tradition, simplified by design to point up specifically what should be counted. The inkblots call forth behavior which retains a high degree of uniquely individual complexity. To look at such behavior and start counting (*M*, *H*, *D*), or testing fruitless hypotheses, is easy. To tease out a pattern, theme or process that constitutes a meaningful unit is a problem that has baffled both clinicians and statisticians. Those clinicians who seem to have the art have had little success in writing the method. Statisticians like R. B. Cattell disdain the "inventive" response of the projective method, throw up their hands, and say: What we want is a traditional model test of dynamisms!

Schafer's approach involves chiefly three elements: (a) a vocabulary or classificatory system taken from the psychoanalytic terminology for ego defenses; (b) the use of judgment in teasing out units, with this judgment based on a background of empirical observation, experimental evidence, and thoughtful speculation; and (c) a rough check back of these broad units against other empirical evidence (descriptive case material). This still crude approach is not new, but Schafer's formulated approximation represents a step forward in specificity and scope.

The author's bias or biases need not be approved with equal enthusiasm. Specific biases that mar some chapters and some elements throughout the book can be mentioned only with an important reservation: as stated here they represent 90 per cent

the reviewer's projections and only 10 per cent the author's attitudinal style. Bluntly, however, on some pages the author does seem to feel that (a) the well-adjusted (analyzed) psychologist in a medical setting should accept the healthy masochistic role of a second-class citizen; (b) an expert is more someone with wide and approved experience than someone who should be asked to produce expert evidence; (c) psychoanalytic theory may not be the ultimate final word, but it is the current final word as far as party line handed down to second-rate citizens is concerned; and (d) usually anything the Rorschach reveals is best understood if labeled with a derogatory word (infantile, sadistic, compulsive).

Actually the author struggles consistently against such biases: lauding solid evidence, rejecting Rorschach's fascinating notes on the inkblots as the final word in this area, and recognizing that there is something "fundamentally neurotic" about reporting all observations in terms of derogatory value judgments. Yet his own problems of professional and scientific identity peek through.

He nevertheless succeeds in setting forth the general outline of a process of clinical judgment, or "intuition," that makes sense. Successful judgments may involve the cancelling out of many details based on false assumptions, self-deception, and initially loose speculation. The relaxed acceptance of all elements, good and bad, checked then with critical rigor against general guide lines which Schafer calls theory, may be important. The general feel for such a process is conveyed by Schafer's book. The further analysis of such judgment processes is important to psychology, as a field of study and as a research tool. The pattern of this approach may lead to more fruitful

progress than a pattern based from the beginning on an unimaginative reading of the APA's *Technical Recommendations for Psychological Tests*.

ROY M. HAMLIN

*Western Psychiatric Institute,
University of Pittsburgh*

TAYLOR, W. S. *Dynamic and abnormal psychology*. New York: American Book Co., 1954. Pp. xiv+658. \$5.50.

According to a statement in the author's preface, this book was designed primarily as a textbook for courses in abnormal psychology. It is my impression, however, that it is unlikely to win a wide acceptance. My reasons for this judgment follow.

It would appear that this book has grown from Professor Taylor's own course in the subject and no doubt will fit it admirably. However, his course seems rather unique, at least so far as my knowledge of other courses and texts in abnormal psychology goes. It is rare that learning, action, thought, and connector processes are discussed in a course in abnormal psychology and treated there much as they would be in general psychology. Yet five of the 19 chapters deal with these topics, and perhaps half the book is devoted to matters not now ordinarily found in currently widely used texts. As a matter of fact, only one chapter is devoted to the major behavioral disorders—neuroses, psychoses, and certain other categories.

More importantly, the several chapters do not seem to hang together in a compellingly coherent way. Having discussed action, for example, the author seems not to make use of this discussion in later treatments of other topics. Moreover, many of the subjects introduced receive so scanty a discussion as to be unintelligible to the naive reader and

simply uninformative to the moderately sophisticated reader. Illustrative, though not representative, is the section entitled, "Psychological Aids to Diagnosis and Prognosis." In two and a fraction pages of actual text, ten topics are discussed, ranging from word association tests and reaction times to projective techniques.

Professor Taylor has obviously read widely, and the book contains numerous quotations from literature and anecdotes of one sort or another drawn from a wide variety of sources. But there is abundance here, to the point often of excessive redundancy, and actual confusion as well as flagging interest is apt to be the result.

Although I feel certain that Professor Taylor has one, the book does not convey to me a theory or a coherent picture of the dynamics of adjustment or of abnormality. It appears to be eclectic, but amorphously so.

It remains to be said that there are many things of value in this volume. The abundant case materials, the many discussions of problems and questions, such, for example, as multiple personality, suggestion, free will, and other matters no longer frequently encountered in our texts, the history of abnormal psychology, and the over-all attempt to employ the categories of general psychology in the special field of dynamics and abnormality are all points of interest and worthy of commendation. As a textbook for general use, however, I do not feel that it can be recommended.

CHARLES N. COFER

University of Maryland

THORPE, LOUIS P., & SCHMULLER, ALLEN M. *Contemporary theories of learning*. New York: Ronald Press Co., 1953. Pp. viii + 480. \$5.50.

The authors indicate that "It is the purpose of this textbook for university and college students to explain the most important theories of learning in the clearest and simplest possible language, to show the relevance of each of them to the educational process, and to point out that in spite of the many conflicts between these theories they have a common ground upon which can be based an intelligible pattern of classroom procedure" (p. v.) Each aspect of this purpose is in itself a large and important undertaking that makes quite different demands upon the authors and requires different evaluation criteria. Two responsibilities are assumed in accomplishing the first part of their purpose: (a) to indicate their criteria for selecting material from a theory, and (b) to explain each theory accurately. As the chapters are written the student is likely to infer that the volume presents unabridged theories of seven men (Thorndike, Guthrie, Hull, Skinner, Wheeler, Tolman and Dewey) and two positions—Functionalism and Gestalt. Since no selection criteria were indicated, the authors are open to the criticism that they have promised more than they have delivered. Also, Thorpe and Schmuller have not always been accurate. For example, they failed to differentiate statements about the theorist's program and metatheory from those pertaining to the theory proper. They inferred from Hull's preoccupation with biological survival and adaptation that these are an integral part of his system. Hull, anticipating confusion on this point, wrote that "adaptive considerations are useful in making a preliminary search for postulates, but that once the postulates have been selected they must stand on their own feet." Adaptation intrudes as a persistent theme in their discussions of most

theories. For example, they say "By learning Guthrie refers to such behavior (acts) as will assist the organism in making necessary adaptations. From this point of view learning is an additive process in that something helpful to him always accrues to the learner as a result of it" (p. 97). This statement also is incorrect, for Guthrie has indicated that the "common-sense" definition of learning as contrasted with a "scientific" one such as his own "applies learning only to the attainment of good results and we shall find that we acquire bad habits and tendencies to failure in exactly the same way in which we acquire good habits and tendencies to success." The authors in purporting to restate Thorndike's law of effect say, "Stated simply, this law holds that the more one utilizes certain neural pathways—assuming that they exist as realities—the stronger become the bonds" (p. 52).

A fundamental difficulty is inherent in the approach taken by the authors to accomplish their second purpose. At the present time it is not possible to apply entire systems of learning to problems because they are not highly developed, logically integrated sets of axioms and postulates. However, application can take place at the level of the more limited special theories, most of which are components of a general theory.

When special theories are applied, three factors appear to determine their utility: (a) the language in which the theory is stated; (b) the reference experiments employed by the theorist; and (c) the particular use to which the theory is to be put. The first two are the responsibility of the theorist; the third is the choice of the practitioner. Both have neglected their responsibilities. Learning theorists have not always provided adequate interpretations of their theo-

retical terms, making it impossible to apply parts of their theory in the sense of deducing relationships that should hold among the elements of any new situation. Those who wish to apply an existing special theory generally have not translated the elements and relationships in the new situation into terms used in stating the theory. This is a necessary prerequisite for the accomplishment of the desired rapprochement between theory and practice. Since this was not done by Thorpe and Schuller, it is not surprising that their final summary of the areas of agreement among learning theorists provides the educator with nothing he did not already know about classroom procedure.

There is another consequence of beginning with well-known "general theories" rather than with a psychological analysis of particular educational problems. Some possible applications were overlooked. Examples include inhibition theory, which could be related to problems of motor skill performance in learning a trade; Gibson's stimulus generalization hypothesis in verbal learning; Cofer and Foley, and also Osgood's mediation hypotheses about language behavior; and Hovland's application of "information theory" to concept learning.

An important oversight is the omission of references to either the tables or figures in the text. Almost none of the 55 figures or 11 tables is referred to in the text.

LAWRENCE M. STOLUROW
University of Illinois

GRUENBERG, SIDONIE M. (Ed.) *The encyclopedia of child care and guidance*. New York: Doubleday, 1954. Pp. 1016. \$7.50.

A most distinguished editorial staff, advisory board, and large roster of contributors were involved

in producing this massive book. Part I, the first two-thirds of the book, contains more than 1,000 alphabetically arranged, cross-referenced entries followed by a classified list of agencies and organizations and an annotated list of further readings. Part II consists of 30 excellent chapters on various aspects of child development and the social forces affecting children.

The encyclopedia portion of the book is consistently addressed to the stereotype of a literate but completely naive parent. The style is chatty and nontechnical while the pages are decorated with an abundance of line drawings, mostly of children. Inasmuch as almost any topic is likely to appear (e.g., Pediatrician, Pediatrics Clinic, Pediculosis Capitis, Peek-A-Boo, Pellagra, Pelvic Examination, Pelvis, Penicillin, Percussion Instruments, Period, Periodicals for Children, Permanent Waves, etc.), the book is guaranteed to furnish even the most professional reader with interesting bits of information. Generally speaking, however, Part II will be of greater interest to psychologists as a popular but comprehensive review of current thinking.

Certainly few will quarrel with the repetitive message: Be loving and patient but seek professional help when a real problem exists. The reviewer foresees little competition for Dr. Spock (himself a contributor) at 35¢ or even Dr. Carmichael at \$12.00.

LEONARD S. KOGAN

*Institute of Welfare Research
Community Service Society of
New York*

SONNEMANN, ULRICH. *Existence and therapy*. New York: Grune & Stratton, 1954. Pp. xi+372. \$5.00.

This will doubtlessly prove to be an exceptionally difficult book for the average psychologist to understand,

whether he be clinician or experimentalist. Not only is Dr. Sonnemann's manner of writing exquisitely involved; but his ideas, aside from the violent emotional reactions they are bound to arouse in most scientifically minded readers, almost defy any purely intellectual comprehension. In Dr. Sonnemann's own existentialist view, these ideas, like all human knowledge, must presumably be experienced-as-being, or known-in-themselves, before they can be factually analyzed or "understood." They must be accepted, as it were, on faith—faith in man's being or existence. This may be so; and certainly logical positivists, who are positively anathema to dyed-in-the-wool existentialists, would never argue that it may not be so, but merely that the question of whether or not it may be so, being essentially unprovable by empirical observation, is a meaningless one. The question is: Why, believing as he does that truth or knowledge can only be directly experienced by man as part of his essential humanity or being, does Dr. Sonnemann (and his fellow existentialists) go to the trouble of long-windedly explaining to his fellow psychologists, in what suspiciously appears to be a highly analytic manner, why existentialism is far superior to every other system of psychology and philosophy ever known to man?

Existence and Therapy is a thoroughgoing discussion of the existentialist viewpoint of Heidegger, Jaspers, Binswanger, Boss, and other recent European philosophers and clinicians. It attempts to make mincemeat of empiricism, objectivism, experimentalism, Freudianism, Gestaltism, Jungianism, and virtually every other influential contemporary way of thinking and therapizing. In so doing, it makes some telling points, particularly in

relation to some of the shortcomings of orthodox Freudians and Gestaltists.

It is on the more positive side that Dr. Sonnemann's detailed presentation of daseinsanalytic theories seems to have much less to offer. Where these theories overlap with neopsychoanalytic and holistic views of man as a total autonomous-autochthonous organism who ceaselessly interacts with other human beings and the external world, and who must be understood and therapized in the light of his biological-social origins and development, existentialism seems eminently sane. But where it goes off, and often, into concepts of being and the naught, existence as being-toward-death, the who of existence, the universality of the love norm, etc., daseinsanalysis appears to become, at least to this empiricism-biased reviewer, distinctly unobjective, mystical, moralistic, and tautological. Its criticisms of other viewpoints sometimes make good sense; but its own espousals often seem to be double-talk.

In any event, Dr. Sonnemann has herewith given us the first quite detailed and reasonably definitive English-language explanation of the daseinsanalytic philosophy. For all its difficulty and blood-pressure raising potentiality, his book is replete with provocative thoughts that certainly deserve a hearing.

ALBERT ELLIS

New York City

PENNINGTON, L. A., & BERG, IRWIN A. (Eds.) *An introduction to clinical psychology*. (2nd Ed.) New York: Ronald Press, 1954. Pp. viii+709. \$6.50.

The student introduced to clinical psychology through this excellent revision of an established text would be

well introduced indeed and he should be both impressed and attracted to further study. He certainly could conclude that clinical psychologists are concerned with many things, from myopia to ethics. He should gain assurance of finding a compatible role in the field whether he is attracted most by the opportunity to carry out research or to apply an art in the interest of humanity. He would, of course, be forewarned that the clinical psychologist is *expected* to burn both ends of the science-service candle, but it will be apparent to him from reading the text that his mentors often burn more brightly at one end than the other.

The involved student might have trouble with Cattell's introductory chapter, particularly if he rereads it after seeing what the other 29 authors have to say. Cattell doesn't seem to know his own mind, and his ideas sometimes seem captive to his phrases. He states that the clinical psychologist should limit his concern to "those who for various reasons have failed in life's educational process" (p. 4), and that "the essential purpose of the profession is to guide the sick back to health" (p. 21). Thus the "ideal practitioner" of clinical psychology should have medical training in addition to his doctoral study in psychology. Finally, for "industrial personnel work, vocational guidance, and educational psychology, the basic and central type of training and qualification needed is that of the clinical psychologist" (p. 20). He is impressed by the derivation of "clinical" from "bed," and it is left to the other authors to disregard etymology and provide operational definitions indicating a much wider scope for clinical psychology. Furthermore the authors of chapters 2 to 25 seem unaware that they are to be saved by factor analysis. A sug-

gestion to the editors in the third edition: keep Cattell, for he is always stimulating, but get two or three others to write introductory chapters entitled "The Meaning of Clinical Psychology."

This second edition is substantially a new book. A hundred pages have been added. Ten new chapters have been included, six have been dropped, and the remaining revised considerably. Notable improvement has been achieved in presenting treatment procedures. The editors have sought unity in a volume likely to fall apart by encouraging the authors to keep problem centered rather than technique centered. Most authors succeed in doing this; the chapters by Shoben, Mowrer, Garner, Dorcus, and Pennington come to mind. Berg gets involved in technique to the point of warning against the distraction in an interview of a cluttered desk or of bright objects in the border of the visual field. Sargent and Hirsch supply a solid chapter on projective methods, possibly more useful to doctoral students reviewing for exams than intriguing to the neophyte. In this edition, a chapter on the professional relationships of the clinical psychologist is replaced by one on research. If the choice is between the two, the focus on research is certainly to be preferred. But it would be useful to the beginning student to know something of the professional problems and responsibilities of the clinical psychologist.

The writing is good throughout the book. Some chapters seem a bit ponderous (like that of Saslow, Guze, and Matarazzo), some simple and highly effective (like that of McCandless), and some brilliant (like that of Mowrer). A great deal of care has gone into the selection of the bibliographies. They should be genuinely useful to the beginning student.

The publishers might have been more generous in their provisions for illustrations, since introductions are better remembered, or at least enjoyed more, when the subject is charming to the eye. Clinical psychology has more to offer visually than visual acuity test charts. Biographical notes on the authors, highlighting their work and thus exhibiting what clinical psychologists do, might add to the value and interest of the book to the beginning student.

Because of the diligence of the editors and the excellence of the authors, clinical psychology now has an even better introductory text than the good one it has had for the past six years.

NICHOLAS HOBBS

Peabody College

KING, H. E. *Psychomotor aspects of mental disease: an experimental study*. Cambridge: Harvard University Press, 1954. Pp. xiv+185. \$3.50.

The current emphasis on projective techniques has unfortunately tended to restrict experimental investigations utilizing other procedures in the study of mental patients. In demonstrating a close relation between severity of behavior disorder and psychomotor performance, as measured by such "old-fashioned" tests as reaction time, speed of tapping, and finger dexterity, King has contributed a quantified technique for measuring the current mental status of the patient. Indirectly, his findings emphasize the importance of the "open" approach in psychodiagnostic work. In the long run progress depends on the development of new techniques rather than on the ultra-refinement of available tools.

JAMES D. PAGE

Temple University

THELEN, HERBERT A. *Dynamics of groups at work*. Chicago: Univer. of Chicago Press, 1954. Pp. x+379. \$6.00

The study of the group has always been a major and is still an increasing concern of social psychology and sociology. And a great many social psychologists (some labeled psychologists and some sociologists) have contributed to what we now know concerning the nature of the human group. Among more recent contributors, Moreno, Lewin, and their followers have been prominent, even though their contributions are not quite as unique and fundamental as their followers persist in claiming.

The intellectual offspring of Lewin include not only social psychologists, who have added to our knowledge of groups and who have shown useful ways to apply this information to the solution of practical problems, but also a growing social engineering cult. This cult has taken and sanctified the words of Lewin and has attributed to him and his followers not only their own substantial findings but also other theories and techniques common among organizational and human relations specialists. These cult members brush up, simplify, and give freshened terminology to such items in order to make them readily saleable to civic leaders, social work directors, education administrators, and business managers. They offer a new dispensation from those mysterious creatures, the "scientific social psychologists." Thelen's book is a new and rather comprehensive exegesis on this cult's materials.

A science finds its application in the form of technologies. Persons functioning as technicians or engineers carry scientific findings into shops and neighborhoods and clinics and board rooms. And as the recent

history of specialists in the various fields demonstrates, engineers quickly seek professional reinforcement in cult-like groups and organizations. Such groups have a pressing "mission" not too well understood and appreciated by the uninitiated or even by the scientists from whom their knowledge derives. They also have initiation requirements and rites, revered leaders, and technical skills dependably applied only by cult members. For the cult to grow sturdily, too, through the increasing production of recognized and recognizable members, its materials must be standardized and rather precisely communicable. As Thelen demonstrates by his book, the group dynamics cult has all these characteristics.

The unique efforts of the scientist flourish best in the hands of individualists, but the fate of technicians and engineers is a joint one. It depends to a large degree upon joint success in jockeying their secular order of specialists into positions of prestige, control, and thus power. In areas where specialties more obviously overlap and thus more sharply compete, such as in human relations, such joint cult-like efforts to cope with competition become all the more mandatory.

This development of a "new" kit of "tools" for management apparently results not only in the caricaturing of tested social psychological theories but also in a striking switch in basic values served. The caricaturing is evinced in Thelen's twelve lists of guidance principles for group workers and groups. These are based far more upon an unquestioning acceptance of a typical middle class societal surrogate's version of current societal morality than upon scientific findings concerning groups. The bare mention of social classes and the fail-

ure to grasp their significance in personality and in group functioning is a part of this problem. The switch in values appears in recurrent emphases by Thelen upon manipulation to achieve morally preconceived goals (rather than upon scientific discovery of the nature of group members, goals, and processes) and upon the service of those concerned with societal stability (rather than upon the stimulation of democratic processes of societal change).

Thelen's first six chapters outline and illustrate six "technologies," which, in his view, while differing, have behind them "fundamental similarities." All six are apparently endorsed. Yet in their outlines, one finds such statements as these: Behavior "is relevant and useful when it contributes to getting the job done, and it is hindering or nonuseful to the extent that it is a response to ideological, class, or racial factors." (Thelen apparently stereotypes "ideological, class, or racial factors"; he does not reveal an appreciation of their implications.) "Neighbors were no longer seen as people one could identify with; the sense of common cause was lost." (He is handicapped by sentimental connotations of the word, neighborhood.) "The role of the student is determined by the will

of the teacher—with some qualification by the standards of the peer and family groups to which the student belongs." (Thelen apparently has not studied how peer and family group participation orients participation in subsequent groups and how class and ethnoic factors alter such participation patterns.)

As I studied this book, I constantly found myself seeing an authoritarian dean of students read it, nod his head in satisfaction, and then go forth and quote formulas from it to justify the disruption of democratic student experiments. When the princes had such as Machiavelli, propaganda analysts for the common man had it difficult enough. Now that they have such polished social engineers as Thelen who sincerely work "for the development of the 'humane community' toward which man's nature . . . is driving him," analysis becomes all the more difficult and efforts to communicate it all the more confusing. But I trust I am overrating the group dynamics cult! Fortunately society has many antitoxins against its continued successful manipulation. Some of them may work slowly, but they work.

ALFRED MCCLUNG LEE

Brooklyn College

EDITORIAL NOTE

As approved by the Board of Directors and the Council of the APA, beginning with January 1956, a new journal will be published by the Association. This journal will review books, monographs, films and related publications—a function presently performed by four different APA journals.

Accordingly, book reviews will appear in the *Psychological Bulletin* only through the completion of the present volume, 52, November 1955 issue.

Hereafter all publications submitted for review and requests to prepare reviews should be directed to the editor of the new journal, *Contem-*

porary Psychology, A Journal of Reviews:

E. G. Boring, Editor
Memorial Hall
Harvard University
Cambridge 38, Mass.

The editors wish to take this opportunity to express their considerable gratitude to the cooperating book reviewers. Any contribution that this department may have made can be attributed only to the work of the reviewers, which we hope has not been a case of love's labour lost.

E. G.
W. D.

PRESENT-DAY PSYCHOLOGY

Edited by A. A. ROBACK

A definitive volume of 40 original contributions embracing practically the whole range of psychology from the neurological basis to military and parapsychology, each chapter written by an expert in his field expressly for this work.

FROM THE TABLE OF CONTENTS

RECENT FINDINGS IN GENERAL NEUROLOGY
Joseph G. Keegan

ISSUES AND RESULTS IN JAPANESE PSYCHOLOGY
P. Ratoosh

THE STATUS OF EMOTION IN CONTEMPORARY
PSYCHOLOGY
Magda Arnold

MINICENTURY PSYCHOBIOGRAPHY
James W. D. Hartman

CLINICAL PSYCHOLOGY
W. G. Ellsberg

RECENT DEVELOPMENTS IN PSYCHOTHERAPY
Emil A. Goshall

INDIVIDUAL PSYCHOLOGY
Rudolf Dreikurs

ABNORMAL PSYCHOLOGY
Philip L. Hartman

SOCIAL PSYCHOLOGY
Ernest L. Galar

APPLIED PSYCHOLOGY
Harold E. Smith

INTRODUCTION TO EXPERIMENTAL PARAPSY-
CHOLOGY
J. B. Rhine

SOME RECENT EXPERIMENTAL WORK IN
PSYCHOBIOGRAPHY
Wynsey Wolff

PRESENT-DAY PSYCHOLOGY IN SWITZER
Emil Frenschke

MILESTONES IN PSYCHOANALYSIS
*Leon J. Saul and
Andrew S. Weissen*

PROJECTIVE TECHNIQUES IN CONTEMPORARY
PSYCHOLOGY
Isaiah Bellak

EDUCATIONAL PSYCHOLOGY
Gordon C. Hansen

ADOLESCENCE
Karl C. Garrison

HYPNOTHERAPY
Milton F. Klein

TRENDS IN STATISTICS AND PROBABILITY IN
PSYCHOLOGY
Herbert E. Sigmund

INTEGRATIONAL PSYCHOLOGY
Clarence Lewis

Approx. 1,000 pages . . . \$12.00

PHILOSOPHICAL LIBRARY, Publishers

12 East 40th St., Desk 22 New York 16, N.Y.

Expedite shipment by prepayment